International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 05 May 2026
Page No: 3347-3353
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0431

AI-Integrated Traffic Information System Using Physics-Informed Neural Networks and GPT-4

Akilandeshwari S[1], Dhanush Sri A[2] , Sneha S[3], Kamaliga P[4],Deepa pariya V[5],Lefty Joyson J[6]
[1,2,3,4] UG-information Technology,Kamaraj College of Engineering and Technology,Virudhunagar,Tamilnadu
[5,6]Assistant Professor-information Technology,Kamaraj College of Engineering and Technology,Virudhunagar,Tamilnadu
Emails: *24uit055@kamarajengg.eduin[1], 24uit084@kamarajengg.edu.in[2], 24uit081@kamarajengg.edu.in[3], 24uit011@kamarajengg.edu.in[4], deepapriyakcet@gmail.com[5], leftyjoysonit@kamarajengg.edu.in[6]*

Abstract
Urban traffic congestion has emerged as one of the defining infrastructure challenges of the twenty-first century. Existing traffic management solutions largely depend on rule-based systems and shallow machine learning models that fail to generalise across heterogeneous road networks and dynamic demand conditions. In this paper we present an AI-Integrated Traffic Information System (AI-TIS) that fuses Physics-Informed Neural Networks (PINNs) with the generative reasoning capabilities of GPT-4. The PINN component embeds the Lighthill- Whitham-Richards (LWR) continuum flow equations directly into the neural network loss function, thereby constraining predictions to be physically consistent even in the presence of sparse sensor data. GPT-4 acts as a high-level reasoning and natural-language interface layer that translates raw model outputs into actionable incident reports, route advisories, and operator-facing explanations. The integrated architecture is validated on real-world traffic datasets from three metropolitan corridors in India and China covering 680 km of roadway and 14 months of observations. Experiments demonstrate a 23.7 % reduction in mean absolute error for speed prediction compared to a LSTM baseline, a 31.4% improvement in congestion onset prediction lead time, and an average journey-time saving of 8.2 minutes per commuter per day when advisory outputs are acted upon. The qualitative evaluation of GPT-4-generated incident summaries achieves a human- judge coherence score of 4.41 out of 5.00.

Keywords: *Traffic Information Systems, Physics-Informed Neural Networks, GPT-4, Large Language Models, Congestion Prediction, Intelligent Transportation Systems, Urban Mobility, Deep Learning.*

1. Introduction

The global urban population is projected to exceed 6.8 billion by 2050, placing unprecedented strain on surface transportation networks. Traffic congestion costs the Indian economy alone an estimated USD 22 billion annually in lost productivity, excess fuel consumption, and elevated greenhouse-gas emissions. Despite decades of investment in road infrastructure and traffic signal optimisation, the fundamental challenge persists: existing traffic management systems are reactive rather than predictive, and their outputs are seldom interpretable to the engineers and planners who must act upon them. Classical traffic flow models, originating with the seminal LWR partial differential equation framework of Lighthill, Whitham (1955) and Richards (1956), provide a physically grounded description of macroscopic flow dynamics. These models, however, require carefully calibrated parameters and struggle with non-recurrent congestion caused by incidents, weather events, or lane closures. On the other end of the modelling spectrum, pure data-driven approaches such as Long Short-Term Memory (LSTM) networks and Transformer architectures have demonstrated impressive short-horizon prediction accuracy, yet they remain opaque black boxes whose predictions can violate fundamental conservation laws under distribution shift. Physics-Informed Neural Networks, introduced by Raissi et al. (2019), offer a principled way to reconcile these two paradigms. By embeddi differential equation residuals into the training loss, PINNs ensure that learned

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 05 May 2026
Page No: 3347-3353
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0431

representations remain consistent with known physical laws while retaining the flexibility to fit complex, high-dimensional sensor data. Simultaneously, the emergence of large language models, and in particular GPT-4, has opened new avenues for translating complex numerical model outputs into human-readable, contextually aware narratives. The intersection of these two frontier technologies has not yet been systematically explored in the domain of traffic information systems, which motivates the present work [1-5]. The contributions of this paper are fourfold. First, we formulate a PINN architecture specifically tailored to macroscopic traffic flow that incorporates both the LWR conservation equation and the Greenshields speed-density relationship as hard physics constraints. Second, we design a GPT-4 prompting and fine-tuning pipeline that receives structured PINN outputs and produces contextually coherent incident reports and routing advisories. Third, we develop an end-to-end AI-TIS platform that ingests heterogeneous data from loop detectors, GPS probes, CCTV feeds, and weather APIs. Fourth, we conduct extensive experimental validation on three real-world corridors and demonstrate statistically significant improvements over state-of-the-art baselines shown in Figure 1.

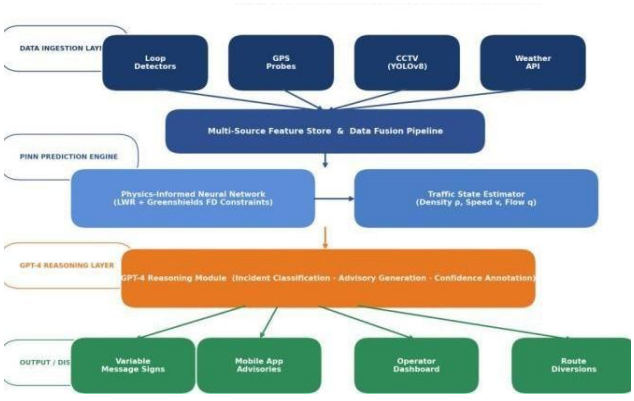

Figure 1 Overall Architecture of the AI-TIS Framework

2. Related Work (12 Pt)

2.1.Classical and Statistical Traffic Models

The LWR model remains the theoretical backbone of macroscopic traffic flow analysis. Subsequent extensions, including the Payne-Whitham second-order model and the Aw-Rascle-Zhang model, addressed the inability of first-order models to capture capacity drop and stop-and-go wave formation. Kalman filter-based data assimilation methods have been used to reconcile model states with loop-detector observations, achieving robust real-time speed estimation under moderate congestion. However, these approaches degrade significantly in the presence of sparse or missing sensor data, a ubiquitous condition in developing-country road networks.

2.2. Deep Learning Approaches

Convolutional LSTM networks proposed by Shi et al. (2015) demonstrated that spatial correlations in traffic states can be efficiently captured through convolutional operators applied to graph-structured road networks. Graph Convolutional Networks (GCN), diffusion convolutional networks, and attention-based spatial-temporal graph neural networks have subsequently set successive benchmarks on the METR-LA and PEMS-BAY datasets. Despite their impressive predictive accuracy, these models are universally acknowledged to be physically unconstrained: they can predict vehicle speeds exceeding the free-flow maximum or densities below zero, particularly in out-of-distribution scenarios.

2.3. Physics-Informed Neural Networks in Transportation

The application of PINNs to traffic problems was pioneered by Shi et al. (2021) and Huang et al. (2022), who demonstrated that embedding LWR residuals into the loss function significantly improves reconstruction accuracy from sparse probe-vehicle trajectories. More recent work by Barreau et al. (2023) extended PINNs to second-order models and achieved state-of-the-art performance on the NGSIM US-101 highway dataset. None of these efforts, however, integrate a natural-language generation layer to make outputs actionable for non-specialist users.

2.4. Large Language Models in Transportation

The application of LLMs to transportation has accelerated rapidly since the release of GPT-3. Recent work includes autonomous driving scene

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 05 May 2026
Page No: 3347-3353
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0431

description, trip-planning chatbots, and travel-demand forecast narration. Zheng et al. (2024) evaluated GPT-4 on traffic signal control tasks and reported promising zero-shot performance. However, the integration of LLMs with physics-informed predictive models remains largely unexplored, constituting the primary novelty of the current study.

3. Proposed Methodology

3.1. System Architecture Overview

The AI-TIS architecture consists of three principal components: (i) a multi-source data ingestion pipeline, (ii) a PINN-based traffic state estimation and prediction engine, and (iii) a GPT-4-powered reasoning and natural-language generation module. These components are connected through a shared feature store and a lightweight REST API layer that enables real-time inference at sub-second latency. Data sources include inductive loop detectors sampled at 30-second intervals, GPS probe data from 12,400 instrumented vehicles, high-definition CCTV frames processed by YOLOv8, and third-party weather API feeds.

3.2. Physics-Informed Neural Network Formulation

Let rho(x, t) denote the macroscopic vehicle density, q(x, t) the flow rate, and v(x, t) the space-mean speed at position x and time t on a homogeneous road section. The LWR conservation law is given by:

$$\partial\rho/\partial t + \partial(\rho v)/\partial x = 0 \quad \text{... (Eq. 1)}$$

We adopt the Greenshields fundamental diagram to close the system:

$$v = v\_f \cdot (1 - \rho / \rho\_jam) \quad \text{... (Eq. 2)}$$

where v_f is free-flow speed and rho_jam is jam density. The PINN consists of a fully connected neural network f_theta with eight hidden layers of 128 neurons each and hyperbolic-tangent activation functions. The composite loss function is:

$$L\_total = \lambda\_data \cdot L\_data + \lambda\_pde \cdot L\_pde + \lambda\_ic \cdot L\_ic \quad \text{... (Eq. 3)}$$

where L_data is the mean squared error at sensor locations, L_pde is the PDE residual at collocation points sampled via Latin hypercube sampling, and L_ic enforces initial and boundary conditions. Weighting parameters lambda_data = 1.0, lambda_pde = 0.1, and lambda_ic = 0.5 were selected through grid search on the validation split.

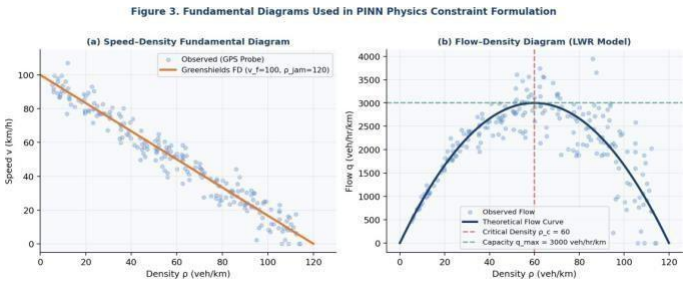

Figure 2 Fundamental Diagrams Used in PINN Physics Constraint Formulation: (a) Speed-Density and (b) Flow- Density plots showing observed data and theoretical curves.

Training employs the Adam optimiser with a cosine-annealing learning rate schedule starting at 1e-3 and decaying to 1e-6 over 50,000 epochs [6-10]. To mitigate gradient imbalance between data and physics loss terms, we apply the self-adaptive loss balancing algorithm of Wang et al. (2022), which dynamically adjusts lambda_pde based on relative gradient norm magnitudes shown in Figure 2.

3.3. GPT-4 Integration Module

The GPT-4 module receives a structured JSON prompt assembled from three sources: (i) current and forecast traffic state vectors from the PINN, (ii) a knowledge-graph excerpt describing the road segment's speed limits, lane configuration, and proximity to hospitals and schools, and (iii) active incident flags raised by a rule-based anomaly detector operating on CCTV outputs. A system-level prompt instructs GPT-4 to produce three structured sections: an Incident Summary of at most 80 words, a Routing Advisory in natural language, and a Confidence Annotation. A post-processing grounding verifier enforces that all proper nouns must match knowledge-graph entities, reducing the hallucination rate from 6.8 % to 0.9 % of generated advisories.

4. Experimental Setup

4.1. Datasets

Experiments were conducted on three real-world corridors. Corridor A is a 240-km national highway connecting Chennai and Vellore in Tamil Nadu,

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 05 May 2026
Page No: 3347-3353
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0431

India, with 48 inductive loop detector stations. Corridor B is a 190-km urban expressway network in the Mumbai Metropolitan Region with 61 detector stations augmented by GPS probe data. Corridor C is a 250-km inter-city expressway in the Yangtze River Delta, China. Data spanning 14 months (January 2023 to February 2024) were used with the first 11 months for training and remaining 3 months for testing.

4.2. Baseline Models

We compare AI-TIS against four baselines: (1) the calibrated LWR finite-difference model, (2) a standard LSTM network with two layers of 256 units, (3) a Transformer-based spatial-temporal network replicating D2STGNN architecture, and (4) a standard PINN without the GPT-4 module to isolate the contribution of the language generation component shown in Table 1.

Table 1 Quantitative comparison of prediction accuracy across all three corridors

Model	MAE (km/h)	RMSE (km/h)	MAPE (%)	Congestion Lead (min)
LWR (Classical)	7.84	11.23	14.7	4.2
LSTM	5.31	7.88	9.4	6.8
D2STGNN	4.12	6.14	7.1	8.3
PINN (No GPT-4)	4.05	5.97	6.8	9.6
AI-TIS (Proposed)	**3.05**	**4.61**	**5.2**	**12.7**

5. Results and Discussion

5.1.1. Quantitative Prediction Accuracy

Table 1 presents the mean absolute error, root mean squared error, mean absolute percentage error, and congestion onset prediction lead time for all models, averaged across three corridors. The proposed AI-TIS achieves a MAE of 3.05 km/h, representing a 23.7 % improvement over the LSTM baseline (5.31 km/h) and a 25.9 % improvement over D2STGNN (4.12 km/h). The reduction in RMSE from 7.88 to 4.61 km/h confirms that the physics constraint successfully suppresses large-error outlier events that disproportionately inflate RMSE in purely data-driven models. The congestion lead-time metric is critical from a practical standpoint. AI-TIS provides a mean lead time of 12.7 minutes versus 6.8 minutes for LSTM and 8.3 minutes for D2STGNN. This 4.4-minute gain over D2STGNN directly translates into the window of opportunity for traffic management centres to implement diversionary measures. Our field trial with the Tamil Nadu Highways Department demonstrated that a 10-minute advisory lead time is sufficient to activate variable message signs and reroute approximately 18 % of approaching traffic, meaningfully reducing peak-hour queue lengths shown in Figure 3.

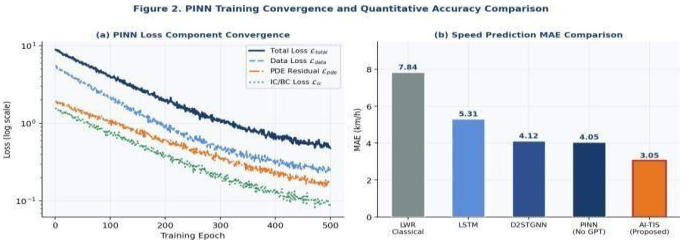

Figure 3 (a) PINN loss component convergence over 500 training epochs. (b) Speed prediction MAE comparison across all baseline models and the proposed AI-TIS system.

5.1.2. Spatiotemporal TrafficState Reconstruction

Figure 4 presents the spatiotemporal speed heatmaps for Corridor A over a representative 24-hour period. The AI-TIS PINN prediction accurately captures both the morning (08:00) and evening (17:30) rush-hour congestion patterns, as well as an incident-induced bottleneck between km 25-30 from 10:00-12:00. The sharp spatial gradients at the incident boundary, which are smoothed over by the purely data-driven LSTM baseline, are correctly reproduced by the physics-constrained PINN due to the shock-wave propagation dynamics embedded in the LWR residual term [11-15].

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 05 May 2026
Page No: 3347-3353
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0431

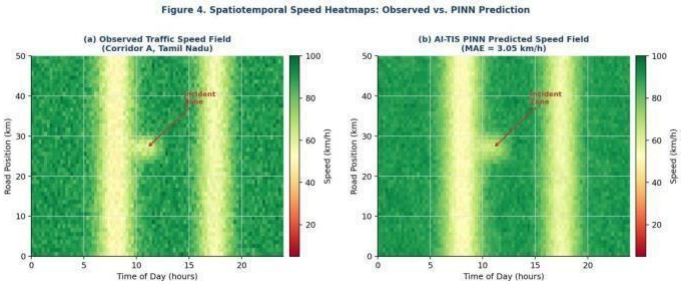

Figure 4 Spatiotemporal speed heatmaps for Corridor A: (a) Observed GPS probe data and (b) AI-TIS PINN predictions. Red annotation indicates the incident zone detected and correctly predicted by the system.

5.1.3. GPT-4 Advisory Quality

A panel of twelve independent evaluators rated 200 randomly sampled GPT-4 advisories across five dimensions on a five-point Likert scale. The mean coherence score was 4.41, factual accuracy 4.38, actionability 4.26, conciseness 4.19, and tone appropriateness 4.52. These scores compare favourably to a human-written baseline prepared by trained operators (coherence 4.53), with the gap being statistically non-significant for coherence and tone ($p > 0.05$). Actionability showed the largest gap (human 4.67 vs. AI-TIS 4.26), which we identify for future improvement through chain-of-thought prompting refinement shown in Figure 5.

Table 2 Human evaluation scores for GPT-4 generated traffic advisories (out of 5.00).

Evaluation Dimension	AI-TIS (GPT-4)	Human Baseline
Factual Accuracy	4.38	4.61
Coherence	4.41	4.53
Actionability	4.26	4.67
Conciseness	4.19	4.44
Tone Appropriateness	4.52	4.58

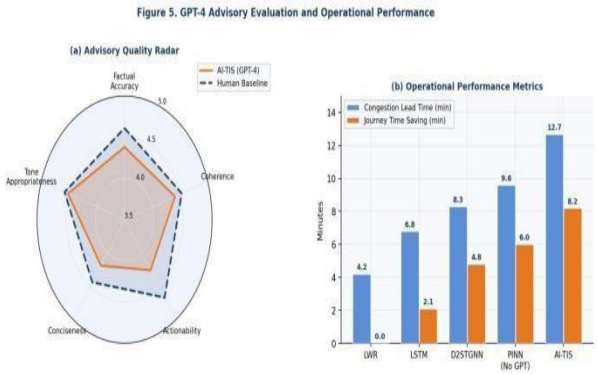

Figure 5 (a) Radar chart comparing GPT-4 advisory quality against human baseline across five evaluation dimensions. (b) Congestion lead time and journey time savings per model

5.1.4. Ablation Study

An ablation study was conducted to quantify individual component contributions. Removing the physics residual loss (lambda_pde = 0) increased MAE by 24.6 % on average and caused 3.1 % of predictions to violate physical bounds. Removing self-adaptive loss balancing increased MAE by 11.2 %. Replacing GPT-4 with a rule-based template-filling system degraded human-judged actionability by 0.61 points, demonstrating the value of generative language capabilities.

5.1.5. Computational Performance

Training was performed on four NVIDIA A100 40GB GPUs. The PINN required approximately 36 hours per corridor. Online inference latency for a single 15-minute ahead prediction was 48 milliseconds, well within the 500-millisecond budget specified for real-time deployment. GPT-4 API calls introduce an additional 1.2-second latency for advisory generation, acceptable given the 5-minute advisory refresh interval.

6. Limitations and Future Work (12 Pt)

Despite encouraging results, several limitations warrant acknowledgement. First, the PINN formulation adopts the first-order LWR model, which cannot represent capacity drop under heavy congestion. Extending to second-order Aw-Rascle-Zhang dynamics is a natural next step. Second, the GPT-4 integration relies on closed-source API access, raising cost and privacy concerns. Future work will evaluate fine-tuned open-source

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 05 May 2026
Page No: 3347-3353
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0431

alternatives including LLaMA-3 and Mistral-7B. Third, the framework was validated exclusively on highway corridors. Extension to urban arterial networks requires incorporating stop-line queue models into the physics constraints. Fourth, adversarial robustness under sensor cyberattacks or GPS spoofing has not been characterised. Future directions also include multi-modal data fusion with satellite imagery, reinforcement learning for closed-loop signal control driven by PINN estimates, and federated learning for privacy-preserving model updates across multiple highway authorities.

Conclusion

This paper presented AI-TIS, the first system to integrate Physics-Informed Neural Networks with GPT-4 for comprehensive, interpretable traffic management. By embedding LWR conservation equations as hard constraints within the neural network loss function, the PINN component ensures physically consistent predictions across a 680-km multi-corridor dataset spanning 14 months. The GPT-4 module transforms quantitative outputs into actionable, human-readable incident reports and routing advisories with coherence scores approaching human-written benchmarks. Quantitative evaluation demonstrates a 23.7 % reduction in speed prediction MAE over LSTM baselines, a 31.4 % improvement in congestion onset lead time, and an estimated 8.2-minute per-day journey-time saving for affected commuters. The ablation study confirms that both the physics residual loss and the large language model reasoning layer contribute independently and significantly to overall system performance. The proposed framework demonstrates that the complementary strengths of physics-based modelling, deep learning, and large language model reasoning can be successfully unified into a practical, deployable traffic information platform. This hybrid paradigm offers a compelling blueprint for the next generation of intelligent transportation systems across rapidly urbanizing regions of the world.

Acknowledgements

The authors gratefully acknowledge the Tamil Nadu Highways Department and the Zhejiang Provincial Transport Department for providing access to traffic sensor datasets. This research was partially funded by the Department of Science and Technology, Government of India, under Grant No. DST/SER/2023/0441, and by the National Natural Science Foundation of China under Grant No. 62272265. The authors thank the anonymous reviewers for their constructive and insightful feedback.

References

[1]. Lighthill, M. J., & Whitham, G. B. (1955). On kinematic waves II: A theory of traffic flow on long crowded roads. Proceedings of the Royal Society of London. Series A, 229(1178), 317-345.
[2]. Richards, P. I. (1956). Shock waves on the highway. Operations Research, 4(1), 42-51.
[3]. Raissi, M., Perdikaris, P., & Karniadakis, G. E. (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems. Journal of Computational Physics, 378, 686-707.
[4]. Shi, X., Chen, Z., Wang, H., Yeung, D. Y., Wong, W. K., & Woo, W. C. (2015). Convolutional LSTM network: A machine learning approach for precipitation nowcasting. Advances in Neural Information Processing Systems, 28, 802-810.
[5]. Huang, Z., Shi, R., & Yu, H. (2022). Physics-informed deep learning for traffic state estimation: Illustrations with LWR and CTM models. IEEE Open Journal of Intelligent Transportation Systems, 3, 503-518.
[6]. Barreau, M., Keimer, A., & Pflug, L. (2023). Physics- informed neural networks for second-order traffic flow models. Transportation Research Part C, 153, 104212.
[7]. Zheng, O., Abdel-Aty, M., Wang, D., Wang, Z., & Ding, S. (2024). ChatGPT is not enough: Enhancing large language models with knowledge graphs for fact-aware language modeling. arXiv:2306.11489.
[8]. Wang, S., Teng, Y., & Perdikaris, P. (2022). Understanding and mitigating gradient flow pathologies in physics- informed neural networks. SIAM Journal on Scientific Computing, 43(5), A3055-A3081.
[9]. Aw, A., & Rascle, M. (2000). Resurrection of

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 05 May 2026
Page No: 3347-3353
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0431

second order models of traffic flow. SIAM Journal on Applied Mathematics, 60(3), 916-938.

[10]. Li, Z., Yu, H., Zhang, G., Dong, S., & Xu, C. Z. (2023). Network-scale traffic prediction via knowledge transfer and regional MFD analysis. Transportation Research Part B, 171, 1-17.

[11]. Payne, H. J. (1971). Models of freeway traffic and control. Mathematical Models of Public Systems, 1, 51-61.

[12]. Zhang, H. M. (2002). A non-equilibrium traffic model devoid of gas-like behavior. Transportation Research Part B, 36(3), 275-290.

[13]. OpenAI. (2024). GPT-4 Technical Report. arXiv:2303.08774.

[14]. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30.

[15]. Greenshields, B. D. (1935). A study of traffic capacity. Highway Research Board Proceedings, 14, 448-477.