COMP8851: Major Project

Session 2, 2024

Final Report

Project Title: Traffic Flow Prediction

Student Name: Ojaswi Kafle

Student ID: 47987731

28 October 2024

Table Of Contents

Abstract

This research project focuses on improving traffic flow prediction by evaluating several predictive models and data sources to fill the gap of reliable real-time traffic prediction and to solve the growing problem of traffic congestion in urban cities. Analyzing over 10+ research papers, it determines the most suitable data sources and models for handling real-time data and other inputs including weather and holiday data. The expected result can offer a clear-cut method for choosing accurate predictive approaches, datasets, and present recommendations for enhancing predictive precision, thus making theoretical and practical enrichments to academic studies and real-world urban traffic management.

1. Introduction

1.1 Aims

This project aims to assess the approaches used for modeling traffic patterns focussing on the factors that might hinder accurate modeling. Its main objectives include:

- Discuss the types and sources of data used by different researchers and evaluate their effectiveness in predictive models (Miglani & Kumar, 2019).
- Evaluate the current approaches in using diverse data in traffic flow prediction models.
- Compare and analyze the performance of the existing traffic prediction models through the various research journals (Sroczyński & Czyżewski, 2023).
- Create strategies for enhancing the accuracy of predicted traffic models.

This initiative is expected to identify the data sources and models that enable accurate traffic flow prediction.

1.2 Background

Traffic congestion in urban areas is becoming more intense worldwide resulting in longer travel times, high consumption of fuel, and emissions of gases in the atmosphere. Forecasting traffic movements is critical to traffic control, viable routing, and congestion avoidance. The latest advances in big data analytics solutions and advanced machine learning approaches create the chance to improve traffic forecast effectiveness and precision (Lv et al., 2014).

The traffic flow problem is dynamic. Researchers from various parts of the world have studied its complexities and have tried different methods to model them accurately. Statistical approaches such as time series analysis and ARIMA (Auto-Regressive Integrated Moving Average) were initially used. These models were found to be accurate during their time in the late 90s and early 2000s, but they were not able to grasp the intricacies of the system (Kashyap et al., 2021). Different factors such as weather, holidays, weekends, road conditions, accidents, and so on affect the traffic flow, and more complicated algorithms were needed. Machine learning models like ANN (Artificial Neural Networks), Random Forests, and SVM (Support Vector Machines) gave better predictions compared to ARIMA; these were better at modeling the dynamic nature of traffic flow. Slowly after mid-2010, neural networks started being mainstream- they could model the system with more accuracy and robustness. These models were modified to hybrid models that coupled different algorithms and hence could utilize the strengths of individual models. Currently, with the evolution of processing power and IoT devices, graph-based methods that attempt to look at the entire picture have been getting popular (Razali et al., 2021).

1.3 Research Problems and Context

The project addresses the limitations and challenges of the existing traffic flow prediction models and suggests improvements regarding data used and modeling methods. The key research problems/questions guiding this study are:

- Are we collecting the right data to make proper forecasts concerning traffic movements?
- What difficulties can be experienced when combining and cleaning datasets?
- What are the algorithms that perform well while predicting traffic flow?
- Do existing models consider unpredictable traffic behavior? If yes, how do they tackle it?

1.4 Expected Outcomes

This research project aims to deliver:

- Comparative Analysis: Comprehensive analysis of various machine learning and predictive methods from past to present in traffic flow prediction.
- Method Validation: Recognition of the most effective traffic prediction methods and the corresponding data sets.
- Insights: In-depth insights into the strengths and limitations of different predictive approaches and datasets used by the researchers.
- Data suitability: Propose the nature and the source of publicly available datasets that can be used for building accurate traffic prediction models.

1.5 Benefits and Significance

This research project significantly impacts traffic flow prediction and urban transportation management in the following ways:

- Foundation for Advanced Traffic Prediction: This sets up a strong foundation for future researchers by highlighting how the right kind of data needs to be chosen and how they can be incorporated into more efficient models for better algorithm decisions.
- Enhancement of Traffic Management: Providing information on dynamic modeling and appraisal of futuristic algorithms that can help resolve traffic challenges in urban areas.
- Environmental and Urban Planning Benefits: The predictions can aid the planning of infrastructure for metropolitan cities by minimizing congestion, and the amount of gas emitted into the atmosphere.
- Facilitation of Modern Technologies: Helps to integrate smart city components and autonomous vehicles into one common traffic system making the entire system efficient and safe.

2. Research Methodology

This research project addresses the problem context of traffic flow forecasting for urban areas. Traffic flow forecasting is essential for the effective control of traffic, organization of transportation, and enhancement of transportation in urban areas. It is therefore determining the right data sets and ‘best bet’ models for predicting traffic flows with reasonable accuracy.

This project's research methodology is divided into three main stages:

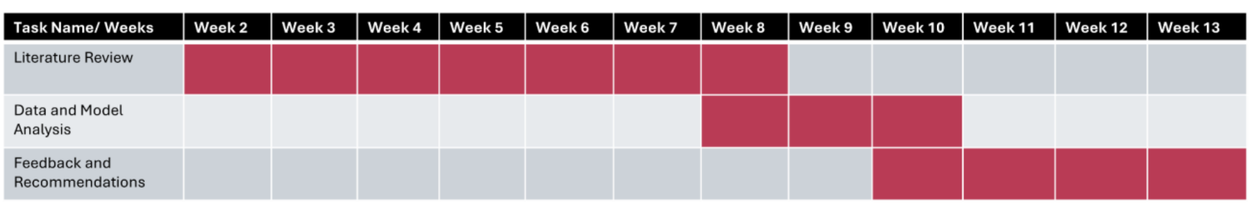
1. **Literature Review:** Conduct a thorough assessment of over ten recent research studies on traffic flow prediction. Examine the models and datasets utilized in these studies. **(Week 2-8)**

2. **Data and Model Analysis:** The research used datasets in the literature to determine the data that can be appropriate for modeling traffic flows. Also, compare different models that the research papers have used for prediction. This will help to evaluate which models would be the best for predicting the traffic flow, what sort of datasets and what type of transformations would do the best job in predicting traffic flow for a given period. **(Week 8-10)**

3. **Feedback and Recommendations:** Summarize the insights from earlier stages and offer suggestions on how to tackle the problem of predicting traffic flow in Sydney. This includes details on optimal dataset model combinations for effective traffic management solutions in future research and real-world use cases to manage urban traffic gridlock in cities, like Sydney. **(Week 10-13)**

The traffic flow prediction problem can be solved effectively by using insights from the literature research, data analysis, and recommendations phases. This research study aims to improve the precision and dependability of traffic flow estimates in metropolitan environments like Sydney by utilizing verified datasets, ideal models, and practical suggestions.

2.1 Project Timeline



3. Literature Review

The field of traffic flow prediction has evolved from traditional statistical methods to advanced machine learning techniques. Individual and comparative studies have been done by many researchers that reveal the advantages and disadvantages of the approaches. In this report, after navigating through a list of algorithms, a premise is created to compare them. There is no one best method and each approach has its shortcomings. This project will also evaluate the available data sources of Sydney and propose how they can be used in different models for better accuracy.

An explanation of modeling techniques used from the past to present is described below:

Traditional Statistical Methods: The time series model divides historical data into trends, seasons, and noise to predict short traffic data in the future. However, it is not so flexible in issues such as emergencies or construction that are bound to affect the road. The next are ARIMA models. They work well for observing short-term trends for small datasets but struggle to handle the nature of diverse live traffic data (Meet Mavani et al., 2023).

Machine Learning Approaches: Support Vector Machines (SVMs) are simple models with minimal learning capacity. We can use them when there's little data to analyze. Random Forests can deal with various types of data and would be useful in informing them of the most critical factors. Also, they are useful for both short-term and long-term predictions (Medina-Salgado et al., 2022).

Deep Learning Techniques: Artificial Neural Networks (ANNs) perform well when it comes to relationships and can consider several factors such as weather or the time of the day. Long Short-Term Memory (LSTM) networks are better at long-term predictions. CNNs (Convolutional Neural Networks) were used in traffic prediction and were found to be great at understanding spatial patterns and working well with LSTMs for a powerful combination (Kashyap et al., 2021).

Hybrid Models: Current hybrid models handle traffic issues due to their capability to incorporate both spatial and temporal data features such as combining CNNs with LSTMs. Such integrations are typically more accurate and much easier for unpredicted and complex urban traffic system environments (Kashyap et al., 2021).

Graph-based Methods: Traffic networks, when assumed to be in the form of graphs, help in estimating the traffic patterns through the large system of roads. They include using Dynamic Spatio-temporal Graph-based CNNs (DST-GCNNs) and Dynamic Multi-Graph Neural Networks (DMGNNs) which are more accurate but more resource intensive (Chen et al., 2020).

Standard error metrics evaluate the models- the mean absolute error (MAE), mean relative error (MRE), and RMS error (RMSE) in all cases.

Several datasets from 2018 have been incorporated into the models through an analysis of the literature review and current research on traffic flow prediction. Here are some of the prominent datasets:

1. **PeMS (Performance Measurement System) datasets:**

These open-source datasets are among the most popular ones in traffic prediction studies. Several subsets of PeMS data are used: PeMS03, PeMS04, PeMS07, and PeMS08.

Datasets	Samples	Nodes	Edges	Time Range
PEMS03	26208	358	547	09/01/2018-11/30/2018
PEMS04	16992	307	340	01/01/2018-02/28/2018
PEMS07	28224	883	866	05/01/2017-08/31/2017
PEMS08	17856	170	295	07/01/2016-08/31/2016

Table 1. PeMS03, PeMS04, PeMS07 and PeMS08 Dataset statistics.

(Source:https://www.researchgate.net/publication/372851257_A_Spatial-Temporal_Feature-Fusion_Model_Based_on_Graph_Convolution_Network_for_Traffic_Flow_Forecasting)

In the figure, four variations of the PeMS dataset are mentioned. Here, the number of detectors is given by the number of nodes. Every detector picks up the traffic volume and vehicle speed every five minutes. The time range is the time interval of which the data is present. Different datasets have different time slices with the traffic data captured every five minutes. The edges give the number of connections the nodes have with each other. In addition to the datasets, there are other variations of PeMS data listed below:

- PeMSD7-M, PeMSD7-L
  - PeMSD7 is a dataset collected from the Caltrans Performance Measurement System (PeMS) by over 39,000 sensor stations in District 7 of California. Data samples from each 30-second interval are aggregated into 5-minute periods.
- METR-LA
  - METR-LA is a traffic speed dataset collected from loop detectors on the LA County Road network. It contains data from 207 selected sensors over 4 months from March to June 2012. The traffic information is recorded at the rate of every 5 minutes, and the total number of time slices is 34,272.
- PeMS-BAY
  - PeMS-BAY is a traffic speed dataset that contains data from 325 sensors in the California Bay Area over 6 months from Jan 1st, 2017, to May 31st, 2017. Here, the traffic information is recorded every 5 minutes, and the total number of time slices is 52,116.

2. **Beijing Traffic Dataset:**
   - BJER4: This dataset originated from the main areas of east ring No.4 routes in Beijing.

3. **Sydney Motorway Traffic Flow dataset**
   - The dataset consists of 208 bi-directional traffic flow counting stations along the M7 Motorway in Sydney, Australia. The dataset is from 2017 and contains 36.34 million data points with traffic data recorded every 3 minutes.

4. **Main Roads Western Australia Historic Data 2013-2024**
   - Data was collected from the Perth Metropolitan State Road Network (PMSRN) at 15-minute intervals. The data has 1021 MLink Identifiers. MLink refers to individual road sections in the metropolitan area. The data is present in the form of CSV files separated by month, downloadable from Western Australia Historic Traffic Data [1].

5. **GPS and cellular phone data:**
   - Live GPS data from automobiles or cellular phone data can be used as additional information that can be fed to the predictions for more accuracy.

The literature review provided shows different approaches used in the research. Among the approaches studied, there are different datasets used and models built with them, and the best predictive models are mentioned. This information about the dataset and best predictive model can then help to discover what are the differences between the datasets. Understanding and a recommendation can be made with this knowledge. It is similar in the case of modeling too- the best model for a type of dataset can be identified and compared.

---

[1] Western Australia Historic Traffic Data: https://bit.ly/historic_trafffic_data_csv

A similar approach of comparing is done by Sayed on AI-based traffic flow prediction (Sayed et al., 2023). This approach, however, will attempt to look for and suggest datasets that can be used (such as Sydney's local data), and recommend a type of model to be created for the best predictions.

4. Results and Discussion

From the study of nine recent research papers on the topic, traffic flow prediction models should use both spatial and temporal features to predict with the greatest accuracy. There were different approaches undertaken in each of the papers, and they were evaluated with simple off-the-shelf models and each of them performed better than them. Different versions of the PEMS dataset were used to train the models and they performed relatively well. The summary of different research papers with the type of models they used, the dataset that they used for their performance compared with other networks, and the peculiarities of the models are shown in the table below.

Research Paper	**Model Name and Type**	**Dataset(s) used**	**Performance**	**Remarks**
A hybrid deep learning-based traffic flow prediction method and its understanding (Wu et al., 2018)	DNN-BTF (Deep Neural Network with Bidirectional Temporal Fusion)	PeMS04	Best performance among Linear, simple Neural network, and Simple autoencoder models.	Graph-based data and graph-based NN that considers spatial and temporal information outperformed other models
LSTM-based traffic flow prediction with missing data (Tian et al., 2018)	LSTM (Long Short-Term Memory)	PeMS 2013 Jan-2013 March and researcher's dataset	Best performance among linear, autoencoder, simple neural network, and support vector models	LSTM is the best model in this case
Spatiotemporal Traffic Flow Prediction with KNN and LSTM (Luo et al., 2019)	KNN-LSTM model	Data from the Transportation Research Data Lab (TDRL) at the University of Minnesota Duluth (UMD) Data Center from March 1st, 2015, to April 30th, 2015	Better results compared with ARIMA, LSTM, SVR (Support Vector Regression), and WVM (Wavelet Neural Network)	Here, the closest point is obtained by KNN and then the closely related stations are modeled

An adaptive hybrid model for short-term urban traffic flow prediction (Hou et al., 2019)	Hybrid combination of ARIMA with WNN with Fuzzy logic	Data of Wenhuadong/Tongyi intersection in Weihai City, China	Hybrid combination performed better than ARIMA and WNN alone which performed better in day and night times respectively	Combining linear and neural networks can extract the advantages of both types of models
Deep and Embedded Learning Approach for Traffic Flow Prediction in Urban Informatics (Zheng et al., 2019)	DELA- consists of an embedding component, a CNN component, and a LSTM component	Anonymized data from the Knowledge Discovery and Data Mining Tools Competition (KDD CUP 2017) and weather data (temperature, air pressure, precipitation, and wind speed)	Better performance compared to ARIMA, LSTM, and RNN models	Complicated model which has a bit of computation but performs better than regulated models
An AutoEncoder and LSTM-Based Traffic Flow Prediction Method (Wei et al., 2019)	**AutoEncoder Long Short-Term Memory (AE-LSTM) prediction model**. It combines the AutoEncoder for feature extraction and LSTM for time series prediction.	Caltrans Performance Measurement System (PeMS)	Demonstrated higher accuracy compared to other methods like CNN and SVM	The hybrid implementation of two networks helps to combine spatial and temporal features and makes the model simpler, thus more efficient
An Effective Spatial-Temporal Attention-Based Neural Network for Traffic Flow Prediction (Do et al., 2019)	**Spatial-Temporal Attention Neural Network (STANN)-** It incorporates both spatial	The dataset used for the experiments is from the **SCATS (Sydney Coordinated Adaptive Traffic System)**	The STANN model outperforms other baseline models, including GRU, LSTM, Seq2Seq,	STANN leverages spatial and temporal attention, allowing it to focus dynamically on the most

	and temporal attention to capture dynamic dependencies in traffic flow data.	in Melbourne, Australia.	DCRNN, and DNN-BTF.	relevant road segments and time steps, resulting in higher accuracy and improved interpretability.
Traffic Flow Prediction Using LSTM with Feature Enhancement (Yang et al., 2019)	LSTM (Long Short-Term Memory) model enhanced with feature engineering techniques that improve the prediction accuracy	PEMS and datasets from local districts of China	The model performed better compared to the BPNN, SVM, SAE, and LSTM models	Feature enhancement was done by a noisy-point smoothing method.
Motorway Traffic Flow Prediction Using Advanced Deep Learning (Mihaita et al., 2019)	CNN, LSTM, CNN-LSTM and BPNN networks	Sydney Motorway Traffic Flow dataset	LSTM outperforms all the models	The LSTM model can be modified using AutoEncoder, or KNN to see if they perform better than standard LSTM

It was good to observe that all the research papers had made some unique transformations in the data and the models to achieve their results. These approaches had their unique impact on the predictive models. It will be beneficial to discuss them in brief.

1. Dataset selection- PEMS data in 5-minute intervals was most evidently used in all the research papers. It may be due to its availability and completeness. Also, most of them attempted to model for only the weekdays and they considered weekends and public holidays to contain high variances.

2. Data preprocessing- For short-term missing data, temporal smoothing was used. For long-term gaps in the data, the average of the previous values was used. Also, in case of high fluctuations, a noisy point smoothing method with feature enhancement was found to be beneficial in the case of LSTM.

3. Modeling- In the case of modeling, variations of LSTM networks are preferred in all cases. This is because the models can identify long-term features as well as adapting to short-term changes in the data. These models can be efficient if the incoming dataset is simplified, eg. by having an autoencoder with LSTM. Or more complex usages like CNN with LSTM networks.

4.1 Suggestion for application to NSW Roads Traffic Volume Data

Sydney's data is available from NSW Roads Traffic Volume Counts API[2]. This API contains hourly data of Sydney, across 208 locations. Some cleaning and transformation need to be done for the data. The data for each hour for a given day is stored in columns. It also has information on whether the date is a public holiday, school holiday, and day of the week which may be helpful for the predictive model.
As this data is on an hourly level, a model can be built with it that makes predictions for the next hour on different road networks. The data being available from 2006 till date should contain all the conditions of the roads.

There should be some trials to be done while modeling this data for better accuracy. One thing that may be beneficial to do may be to sample the hourly data into 5-minute interval data and then use this for predictive modeling. This will help to forecast the data for the next N-5-minute intervals, ie. traffic volume for the next 5, 10, 15, and 20 minutes.

For predictive modeling, it will be best to start with a base LSTM model and base CNN first for the dataset. Then different versions of them can be used to experiment to find the model that best fits the dataset. More complex models such as the CNN-LSTM model, AE-LSTM model, and KNN-LSTM models can be created as well so that the error decreases.

Alternatively, an experiment can be done where an LSTM model is trained on the Main Roads Western Australia dataset in 15-minute intervals. This approach may help to grasp the road traffic changes in 15 minutes supposing that Sydney and Perth have similar traffic patterns. The resulting model can be fine-tuned using Sydney's dataset.

4.2 Feedback for Future Research

Combining traffic flow data with external datasets, including calendar data, weather data from the Bureau of Meteorology (BOM) weather datasets[3], GPS data from cars, and roadwork and closure details, accuracy in traffic flow predictions may be much improved. The volume of traffic during different times of day, weeks, months, and even seasons often fluctuates. Usually, traffic is down on public holidays (or at least down relative to weekdays), but they can spike at very abnormal times. Using binary indicators or categorical data for these calendar occurrences allows one to better predict the temporal subtleties of traffic flow.

The real-time data on speed and congestion of car traffic derived from car GPS data makes possible more dynamic traffic projections. Incidentally, this implies that such data can be useful to predict possible driving behaviors under different conditions. As road construction operations often lead to significant detours and delays, integration of road work and closure data is necessary. Using predictive models, it is possible to achieve a greater degree of analysis of the state of urban traffic by taking into consideration the impact of ongoing construction works together with meteorological and calendar data. With the help of modeling techniques like LSTMs or gradient boosting algorithms, accuracy can be further improved due to these large datasets. This comprehensive approach enhances the capacity for forecasting at the same time as it provides appropriate information to traffic management authorities and municipal planners, which will eventually lead to a more efficient transport system in Sydney.

---

[2] NSW Roads Traffic Volume Counts API: https://data.nsw.gov.au/data/dataset/2-nsw-roads-traffic-volume-counts-api/resource/d1180fec-7143-4171-9397-978a249339c7

[3] Bureau of Meteorology (BOM) weather datasets: http://www.bom.gov.au/

5. Conclusion and Future Outlook

This research project analyses traffic flow prediction in a systematic way from a methodological three-phase process of grasping different types of datasets and models from recent research journals. This approach enhanced my understanding of various traffic scenarios in different cities. Specifically, this project has refined my knowledge regarding how traditional models like ARIMA were used in the past and how they are being replaced by better deep-learning hybrid models. It was also insightful to notice that simple models like KNN and Auto Encoders can be used to do a first level of modeling before passing the data to a more complex LSTM model. For traffic flow prediction, LSTM models are found to be the most accurate.

Similarly, the project identifies that in most of the research papers, the PEMS dataset and its variations were used, which contains data on vehicles every five minutes to fifteen minutes. Looking at the nature of the source data and the methods of cleaning that the research papers have undertaken, my recommendations on how to use the publicly available Sydney traffic data to model traffic flow are presented. If a predictive model is to be built for Sydney, the best data to use would be Sydney's traffic dataset as it will have the trends of this region. For example, instances like the Melbourne Cup cannot be modeled by the other datasets. Also, the weather, calendar data, and possibly road closure datasets would prove to be good input variables to analyze the trends of traffic flow. These will make predictions more accurate.

Looking forward, several challenges should be addressed in future research, including the feasibility, modularity, and real-time accuracy of various predictive models for numerous urban contexts. Real-time adaptation mechanisms, where a solution can learn from changes that may happen in the real world, e.g., changes in weather or traffic conditions can be provided as input to the models. Also, transfer learning, where a solution used in one context is applied to another related context can be crucial if the training dataset size is small. For modeling Sydney's traffic, it would make sense to use Sydney's traffic data as a primary source of data, and the recommendations are provided in previous sections of the project. This ability to predict traffic flow could bring system improvements in managing traffic in Sydney. It will ultimately help to make the road traffic network more resilient and efficient.

References

Chen, K., Chen, F., Lai, B., Jin, Z., Liu, Y., Li, K., Wei, L., Wang, P., Tang, Y., Huang, J., & Hua, X.-S. (2020). Dynamic Spatio-Temporal Graph-Based CNNs for Traffic Flow Prediction. *IEEE Access*, *8*, 185136–185145. https://doi.org/10.1109/access.2020.3027375

Do, L. N. N., Vu, H. L., Vo, B. Q., Liu, Z., & Phung, D. (2019). An effective spatial-temporal attention based neural network for traffic flow prediction. *Transportation Research Part C: Emerging Technologies*, *108*, 12–28. https://doi.org/10.1016/j.trc.2019.09.008

Hou, Q., Leng, J., Ma, G., Liu, W., & Cheng, Y. (2019). An adaptive hybrid model for short-term urban traffic flow prediction. *Physica A: Statistical Mechanics and Its Applications*, *527*, 121065. https://doi.org/10.1016/j.physa.2019.121065

Kashyap, A. A., Raviraj, S., Devarakonda, A., Nayak K, S. R., K V, S., & Bhat, S. J. (2021). Traffic flow prediction models – A review of deep learning techniques. *Cogent Engineering*, *9*(1). https://doi.org/10.1080/23311916.2021.2010510

Luo, X., Li, D., Yang, Y., & Zhang, S. (2019). Spatiotemporal Traffic Flow Prediction with KNN and LSTM. *Journal of Advanced Transportation*, *2019*, 1–10. https://doi.org/10.1155/2019/4145353

Lv, Y., Duan, Y., Kang, W., Li, Z., & Wang, F.-Y. (2014). Traffic Flow Prediction with Big Data: A Deep Learning Approach. *IEEE Transactions on Intelligent Transportation Systems*, *16*(2), 1–9. https://doi.org/10.1109/tits.2014.2345663

Medina-Salgado, B., Sánchez-DelaCruz, E., Pozos-Parra, P., & Sierra, J. E. (2022). Urban traffic flow prediction techniques: A review. *Sustainable Computing: Informatics and Systems*, *35*, 100739. https://doi.org/10.1016/j.suscom.2022.100739

Meet Mavani, Chaturvedi, P., & Swati Manekar. (2023). Comparative Analysis of Time Series Models for Traffic Flow Prediction. *2023 10th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering*, *10*. https://doi.org/10.1109/upcon59197.2023.10434410

Miglani, A., & Kumar, N. (2019). Deep learning models for traffic flow prediction in autonomous vehicles: A review, solutions, and challenges. *Vehicular Communications*, *20*, 100184. https://doi.org/10.1016/j.vehcom.2019.100184

Mihaita, A.-S., Li, H., He, Z., & Rizoiu, M.-A. (2019). Motorway Traffic Flow Prediction using Advanced Deep Learning. *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*. https://doi.org/10.1109/itsc.2019.8916852

Razali, N. A. M., Shamsaimon, N., Ishak, K. K., Ramli, S., Amran, M. F. M., & Sukardi, S. (2021). Gap, techniques and evaluation: traffic flow prediction using machine learning and deep learning. *Journal of Big Data*, *8*(1). https://doi.org/10.1186/s40537-021-00542-7

Sayed, S. A., Abdel-Hamid, Y., & Hefny, H. A. (2023). Artificial intelligence-based traffic flow prediction: a comprehensive review. *Journal of Electrical Systems and Information Technology*, *10*(1). https://doi.org/10.1186/s43067-023-00081-6

Shen, H., & He, M. (Eds.). (2023). *PeMS03, PeMS04, PeMS07 and PeMS08 Dataset statistics*. https://www.researchgate.net/publication/372851257_A_Spatial-Temporal_Feature-Fusion_Model_Based_on_Graph_Convolution_Network_for_Traffic_Flow_Forecasting

Sroczyński, A., & Czyżewski, A. (2023). Road traffic can be predicted by machine learning equally effectively as by complex microscopic model. *Scientific Reports*, *13*(1). https://doi.org/10.1038/s41598-023-41902-y

Tian, Y., Zhang, K., Li, J., Lin, X., & Yang, B. (2018). LSTM-based traffic flow prediction with missing data. *Neurocomputing*, *318*, 297–305. https://doi.org/10.1016/j.neucom.2018.08.067

Wei, W., Wu, H., & Ma, H. (2019). An AutoEncoder and LSTM-Based Traffic Flow Prediction Method. *Sensors*, *19*(13), 2946. https://doi.org/10.3390/s19132946

Wu, Y., Tan, H., Qin, L., Ran, B., & Jiang, Z. (2018). A hybrid deep learning based traffic flow prediction method and its understanding. *Transportation Research Part C: Emerging Technologies*, *90*, 166–180. https://doi.org/10.1016/j.trc.2018.03.001

Yang, B., Sun, S., Li, J., Lin, X., & Tian, Y. (2019). Traffic flow prediction using LSTM with feature enhancement. *Neurocomputing*, *332*, 320–327. https://doi.org/10.1016/j.neucom.2018.12.016

Zheng, Z., Yang, Y., Liu, J., Dai, H.-N., & Zhang, Y. (2019). Deep and Embedded Learning Approach for Traffic Flow Prediction in Urban Informatics. *IEEE Transactions on Intelligent Transportation Systems*, *20*(10), 3927–3939. https://doi.org/10.1109/tits.2019.2909904