

A Survey on Vehicular Traffic Flow Anomaly Detection Using Machine Learning

Jackel Vui Lung Chew^{1} and Mohammad Fadhli Asli¹*

¹Optimisation and Visual Analytics Research Lab, Faculty of Computing and Informatics, Universiti Malaysia Sabah Labuan International Campus, 87000, Labuan, Malaysia

Abstract. Vehicular traffic flow anomaly detection is crucial for traffic management, public safety, and transportation efficiency. It assists experts in responding promptly to abnormal traffic conditions and making decisions to improve the traffic flow. This survey paper offers an overview of the application of machine learning to detect anomalies in the traffic flow. Through an extensive review of the literature from the Scopus database, this paper explores the technical aspects of traffic flow anomaly detection using machine learning, including data sources, data processing approaches, machine learning algorithms, and evaluation metrics. Additionally, the paper highlights the emerging research opportunities for researchers in enhancing traffic flow anomaly detection using machine learning.

1 Introduction

Vehicular traffic flow anomaly is an unusual behaviour observed in the flow of vehicles on the road or in a traffic system. This anomaly can be referred to as phantom traffic jams in which traffic congestion and slowdowns occur on the road for no apparent reason. This anomaly is not associated with collisions, obstructions, or lane closures [1]. Traffic flow anomalies are persistent challenges to the traffic management authorities. For instance, managing traffic flow requires costly technologies to implement and maintain [2,3]. Then, congestion in traffic causes economic losses due to wasted fuel and time delay [4,5]. In addition, congested traffic increases greenhouse gas emissions, contributing to air pollution and climate change [6,7]. Moreover, irregular traffic flow increases the likelihood of traffic accidents [8-10].

Detecting anomalies in traffic flow is essential for appropriate responses to traffic incidents and optimising traffic management approaches. Machine learning (ML) offers a promising solution to tackle traffic flow anomaly detection challenges. For example, ML can analyse historical traffic data to predict various traffic conditions. These predictions can help traffic management authorities make informed routing to optimise traffic flow [11,12]. Predictions made by ML can also help reduce economic losses by reducing congestion and optimising traffic flow [13]. Furthermore, ML can be utilised to develop analytical models for classifying accident-prone locations and times. These models allow relevant authorities to implement appropriate safety measures [14] proactively.

* Corresponding author: jackelchew93@ums.edu.my

Several authors have provided reviews or surveys on the application of ML in vehicular traffic flow. However, it was found that limited literature mainly focuses on detecting traffic flow anomalies. To the best of our search of existing surveys and reviews about traffic flow anomaly detection using ML, two published survey and review papers partially like our paper [15,16]. The limitations of these papers have been identified and can be explained as follows. The former surveyed ML in optimising operations in freight transportation, supply chain, and logistics management. A brief review of anomaly detection on transportation data using ML is provided. However, the review of the technical aspect of ML to detect anomalies in transportation data is limited. The latter comprehensively discussed ML methods used in intelligence transportation systems applications. However, the literature gap in the technical aspect of ML to detect anomalies in traffic flow still needs to be addressed.

This survey paper aims to provide an overview of the application of ML to detect anomalies in traffic flow. This paper focuses on the technical aspects of ML in developing a traffic anomaly detection mechanism. This paper also highlights the forward-looking research opportunities of using ML that can enhance the technologies in traffic anomaly detection mechanisms. The main contributions of this paper are to reveal the data source, data processing, algorithm selection, and model evaluations used by recent research and to provide emerging research opportunities from a different point of view that focuses on the technical aspect of ML applications. These two contributions make this paper different from the existing surveys and reviews [15,16].

This survey adopted the PRISMA guidelines [17]: Firstly, we defined paper inclusion criteria such as year range from 2019 until the present, language must be in English only, and publication status is published. Secondly, we formed a search entry for the Scopus database. We defined the following search expression: (TITLE-ABS-KEY (“vehicular” OR “vehicle”) AND TITLE-ABS-KEY (“traffic”) AND TITLE-ABS-KEY (“anomaly”) AND TITLE-ABS-KEY (“machine learning”)). We excluded grey literature, duplicates, and pre-prints from the publications potentially relevant to this survey focus. Then, we examined the publications' abstracts, titles, and keywords to identify papers that consider ML development for traffic flow anomaly detection. Thirdly, we defined paper exclusion criteria such as not related to the vehicular traffic flow anomaly, not review and survey papers, having no detailed report about ML technical aspects and having no full text available. After carefully examining the eligible papers, we also extended the list of identified papers to relevant ones in the inner references.

This paper is organised as follows. Section 2 presents the technical aspects of ML in traffic flow anomaly detection mechanisms. Section 3 then highlights the emerging research opportunities on this topic. Section 4 concludes the paper.

2 Traffic Flow Anomaly Detection Using Machine Learning

Generally, developing a traffic flow anomaly detection mechanism using ML involves vital processes such as data source, data processing, algorithm selection, and model evaluations. Only some other technical details related to ML for detecting traffic flow anomalies are included to ensure only advanced and recent developments are considered for the survey and target readers gain sufficient details about the technical aspects of ML. This section presents the data sources, data processing approaches, ML algorithms, and evaluation metrics used by the selected literature.

2.1 Data Source

In the context of ML, the data source refers to the origin of the data gathered for training and testing the ML. Data is a fundamental component of ML, and understanding data sources is

essential in developing effective ML. Table 1 shows the list of studies on the type of data sources used to develop ML for traffic flow anomaly detection.

Table 1. Data sources used by the selected studies.

Authors	Type of Data Sources
Wang et al. [18]	Simulation of Urban Mobility and Historical Traffic Datasets
Abbas et al. [19]	Next-Generation Simulation Datasets
Peng et al. [20]	Historical Traffic and Publicly Available Datasets
Sabour et al. [21]	Simulation of Urban Mobility and Video Datasets
Ma et al. [22]	Publicly Available Datasets
Meena et al. [23]	Video Datasets
Makhmutova et al. [24]	Video Datasets
Mon et al. [25]	Simulation of Urban Mobility Datasets

Based on Table 1, most of the literature reportedly used Simulation of Urban Mobility (SUMO) and video datasets as their data sources. SUMO is an open-source traffic simulation package designed to support the simulation and analysis of urban traffic systems. Researchers can utilise it to model ML and analyse traffic flow anomalies in realistic and complex urban environments. These datasets typically include information about road networks, traffic flow, vehicle types, routes, and other relevant parameters. Thus, researchers can generate realistic traffic scenarios, simulate diverse traffic conditions, and customise desired traffic environments. On the other hand, video provides a continuous stream of real-time visual traffic data that includes traffic patterns, vehicle movements, and environmental conditions on the road. Researchers can utilise video data in developing ML models and leverage computer vision techniques to extract valuable features such as vehicle speed, trajectory, lane changes, and unusual behaviours.

2.2 Data Processing Approach

Data processing refers to the set of activities to prepare raw data before it can be fed to ML for training, validation, and testing. The data processing approaches are different according to the type of data. Table 2 shows the list of studies with the data processing activities used in their studies.

Table 2. Data processing approaches used by the selected studies.

Authors	Data Processing Approach
Wang et al. [18]	Data labelling
Abbas et al. [19]	Data exploration and analysis, data preprocessing, data splitting, data normalisation, feature learning, outlier detection
Peng et al. [20]	Data splitting, feature learning
Sabour et al. [21]	Data splitting
Ma et al. [22]	Data preprocessing

Meena et al. [23]	Data preprocessing, feature learning, data splitting
Makhmutova et al. [24]	Data preprocessing
Mon et al. [25]	Data splitting

Based on Table 2, different studies used different sets of data processing activities. By omitting the order of the activities by the authors, this paper discusses common key activities: data splitting, data preprocessing, and feature learning. Data splitting is an activity in ML where the dataset is divided into several subsets for training and testing. The training set is used to train the ML, while the testing set is used to evaluate the ML performance. On the other hand, data preprocessing is used to ensure that the data is meaningful and essential for optimising the performance of an ML. Besides these two activities, feature learning is used for discovering and extracting meaningful features from the data. Feature learning allows an ML to discover these features during the ML model training.

2.3 Machine Learning Algorithm

ML algorithm is a set of rules for a computer to learn patterns from the data and make predictions. The algorithm determines how the ML learns and generalises the training data into new data, which can be considered predictions. Table 3 shows the list of studies with the proposed ML algorithms in their studies.

Table 3. Machine learning algorithms used by the selected studies.

Authors	Proposed Algorithm	Type of ML
Wang et al. [18]	Graph Neural Network with Multi-instance Learning Long Short-Term Memory with Multi-instance Learning	Supervised
Abbas et al. [19]	Combination of Deep Learning, Autoencoders, and Z-score Method	Supervised and Unsupervised
Peng et al. [20]	Combination of One-Class Support Vector Machine and Informer Method	Unsupervised
Sabour et al. [21]	DeepFlow based on the Siamese Network	Semi-Supervised
Ma et al. [22]	Gaussian Mixture Model, Partial Least Squares Regression, Wavelet Transform, and Support Vector Regression	Supervised and Unsupervised
Meena et al. [23]	Convolutional Long Short-Term Memory with K means	Unsupervised
Makhmutova et al. [24]	Region-based Convolutional Neural Network with Inception-ResNet	Unsupervised
Mon et al. [25]	Support Vector Machine with Radial Basis Function	Supervised

Based on Table 3, this paper found that the proposed ML algorithms for traffic flow anomaly detection involve combining two or more ML techniques. Moreover, the distinction between the types of ML depends on the type of learning by the authors and the need for training data labelling. The supervised ML authors trained the labelled data before making predictions or classifications. Then, the authors who used unsupervised ML explored the structure of unlabelled data before discovering patterns or relationships. One selected paper used a semi-supervised ML combining labelled and unlabelled data.

2.4 Evaluation Metric

In the context of ML, evaluation metrics are the measures used to assess the performance of an ML. These metrics determine the performance of an ML by comparing the same metrics with different MLs. The choice of the metrics depends on the specified task of the ML. Table 4 shows the list of studies with the metrics used in their studies.

Table 4. Evaluation metrics used by the selected studies.

Authors	Evaluation Metrics
Wang et al. [18]	Area Under Receiver Operating Characteristic and Area Under Precision-Recall Curves
Abbas et al. [19]	Accuracy
Peng et al. [20]	Precision, Recall and F1 Score
Sabour et al. [21]	Receiver Operating Characteristic Curve, Precision-Recall Curve, Precision, Recall and F1 Score
Ma et al. [22]	Precision, Recall and F1 Score
Meena et al. [23]	Accuracy
Makhmutova et al. [24]	Accuracy
Mon et al. [25]	Accuracy, Sensitivity, Specificity, Precision, Recall and F1 Score

Based on Table 4, most authors used ML evaluation metrics such as Accuracy, Precision, Recall and F1 Score. These metrics are used to evaluate the performance of classification-type ML and are helpful when dealing with imbalanced datasets.

3 Emerging Research Opportunities

Based on the extensive review of the selected literature shown in Section 2, this survey has discovered several emerging research opportunities that future researchers and practitioners can pursue in ML technology:

- More realistic and complex publicly available real-world datasets of traffic flow should be added to the literature as the data sources. High dependent on simulation datasets might lead to a saturation stage of developing effective ML.
- Traffic flow anomaly detection that can dynamically adapt to environmental factors and road conditions changes has gained much interest. Thus, standard data processing approaches for different data types should be established.
- Integrating multiple basic and hybrid ML should be explored further to enhance the robustness and interpretability of traffic flow anomaly detection mechanisms.
- Benchmarks and standards for evaluating the performance of traffic flow anomaly detection using ML should be established so that comparisons between different approaches and ML techniques can be reasonably made.

4 Conclusion

This paper has presented a survey on the application of ML to detect anomalies in vehicular traffic flow, focusing on its technical aspects. The survey reviewed the data sources, data processing approaches, ML algorithms, and evaluation metrics reported by relevant ML works. This paper found that SUMO datasets are widely utilised as a data source that enables

realistic simulations of complex urban traffic environments. Then, this paper highlighted common data processing activities such as data splitting, feature learning, and data preprocessing. ML algorithms proposed in the surveyed literature ranged from supervised to unsupervised and semi-supervised, often combining multiple techniques to enhance anomaly detection. In addition, evaluation metrics used for assessing ML performance included precision, recall, and F1 score. This paper also identified emerging research opportunities, urging the incorporation of realistic, real-world traffic datasets, the development of adaptable anomaly detection systems, the exploration of hybrid ML approaches, and the establishment of benchmarks for fair performance evaluations. The significance of this paper in addressing the literature gap related to detecting traffic flow anomalies using ML is to lead the development of more accurate ML models for identifying and classifying traffic anomalies. Valuable insights that can be obtained from ML models can assist decision-makers in managing traffic, ensuring public safety, and enhancing city transportation efficiency. This survey can guide future research efforts for advancements in ML technology that can improve traffic management, public safety, and city transportation efficiency.

ACKNOWLEDGEMENT

Universiti Malaysia Sabah supports the research under the university research grant SBK0462-2021 received by the first author.

References

1. K. Kalair, C. Connaughton, *Transp. Res. Part C Emerg. Technol.* **127**, 103178 (2021)
2. R. Ravish, S. R. Swamy, *Transp. Telecommun.* **22**, 163-182 (2021)
3. M. W. Ei Leen, N. H. A. Jafry, N. M. Salleh, H. Hwang, N. A. Jalil, *Mitigating Traffic Congestion in Smart and Sustainable Cities Using Machine Learning: A Review*. in Proceedings of International Conference on Computational Science and Its Applications, ICCSA, 3-6 July 2023, Athens, Greece (2023)
4. M. A. AlAttar, N. Z. Al-Mutairi, *Int. J. Crashworthiness* **26**, 258-269 (2021)
5. M. A. Fattah, S. R. Morshed, A. A. Kafy, *Transp. Eng.* **9**, 100122 (2022)
6. J. Lu, B. Li, H. Li, A. Al-Barakani, *Cities* **108**, 102974 (2021)
7. N. Abdull, M. Yoneda, Y. Shimada, *Air Qual. Atmos. Health* **13**, 731-738 (2020)
8. H. Chang, L. Li, J. Huang, Q. Zhang, K. S. Chin, *Accid. Anal. Prev.* **169**, 106618 (2022)
9. S. Sánchez González, F. Bedoya-Maya, A. Calatayud, *Sustainability* **13**, 7500 (2021)
10. D. Albalate, X. Fageda, *Transp. Policy* **105**, 145-152 (2021)
11. S. J. Kamble, M. R. Kounte, *Procedia Comput. Sci.* **171**, 2235-2241 (2020)
12. A. Boukerche, J. Wang, *Comput. Netw.* **181**, 107530 (2020)
13. R. de la Torre, C. G. Corlu, J. Faulin, B. S. Onggo, A. A. Juan, *Sustainability* **13**, 1551 (2021)
14. Y. Berhanu, E. Alemayehu, D. Schröder, *J. Adv. Transp.* **2023**, 6643412 (2023)
15. E. Dilek, M. Dener, *Sensors* **23**, 2938 (2023)
16. K. Tsolaki, T. Vafeiadis, A. Nizamis, D. Ioannidis, D. Tzovaras, *ICT Express* **9**, 284-295 (2023)
17. A. Liberati, D. G. Altman, J. Tetzlaff, C. Mulrow, P. C. Gøtzsche, J. P. Ioannidis, M. Clarke, P. J. Devereaux, J. Kleijnen, D. Moher, *J. Clin. Epidemiol.* **62**, e1-e34 (2009)
18. Z. Wang, Z. Li, Z. Li, Y. Xu, F. Qi, J. Kong, *Comput. Secur.* **132**, 103362 (2023)

19. S. Abbas, M. O. Malik, A. R. Javed, S. P. Hong, *Electronics* **12**, 2072 (2023)
20. X. Peng, Y. Lin, Q. Cao, Y. Cen, H. Zhuang, Z. Lin, *Traffic Anomaly Detection in Intelligent Transport Applications with Time Series Data using Informer*. in Proceedings of IEEE 25th International Conference on Intelligent Transportation Systems, ITSC, 8-12 October 2022, Macau, China (2022)
21. S. Sabour, S. Rao, M. Ghaderi, *DeepFlow: Abnormal Traffic Flow Detection Using Siamese Networks*. in Proceedings of IEEE International Smart Cities Conference, ISC2, 7-10 September 2021, Manchester, United Kingdom (2021)
22. Y. Ma, Z. Zhang, S. Chen, Y. Yu, K. Tang, *IEEE Access* **7**, 8028-8038 (2019)
23. K. Meena, A. Viji, J. J. Athanesious, V. Vaidehi, *Detecting Abnormal Event in Traffic Scenes using Unsupervised Deep Learning Approach*. in Proceedings of International Conference on Wireless Communications Signal Processing and Networking, WiSPNET, 24-26 March 2022, Chennai, India (2019)
24. A. Makhmutova, R. Minnikhanov, M. Dagaeva, I. Anikin, T. Bolshakov, I. Khuziakhmetov, *Intelligent Detection of Object's Anomalies for Road Surveillance Cameras*. in Proceedings of International Multi-Conference on Engineering, Computer and Information Sciences, SIBIRCON, 21-22 October 2019, Novosibirsk, Russia (2019)
25. E. E. Mon, H. Ochiai, C. Saivichit, C. Aswakul, *Traffic Anomaly Classification by Support Vector Machine with Radial Basis Function on Chula-SSS Urban Road Network*. in Proceedings of 9th International Workshop on Computer Science and Engineering, WCSE, 15-17 June 2019, Hong Kong, China (2019)