sustainability

MDPI

Article

Towards Green Innovation in Smart Cities: Leveraging Traffic Flow Prediction with Machine Learning Algorithms for Sustainable Transportation Systems

Xingyu Tao [1], Lan Cheng [2,*], Ruihan Zhang [3], W. K. Chan [4], Huang Chao [5] and Jing Qin [1]

1 Centre for Smart Health, School of Nursing, The Hong Kong Polytechnic University, Hong Kong, China; xingyu26.tao@polyu.edu.hk (X.T.); harry.qin@polyu.edu.hk (J.Q.)
2 Big Data Bio-Intelligence Laboratory, Big Data Institute, The Hong Kong University of Science and Technology, Hong Kong, China
3 The Department of Computing, The Hong Kong Polytechnic University, Hong Kong, China; rzhangao@connect.ust.hk
4 The Department of Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong, China; wkinchan@outlook.com
5 School of Arts and Design, Shenzhen University, Shenzhen 518060, China; chao@szu.edu.cn
* Correspondence: chenglan@ust.hk

Abstract: The emergence of smart cities has presented the prospect of transforming urban transportation systems into more sustainable and environmentally friendly entities. A pivotal facet of achieving this transformation lies in the efficient management of traffic flow. This paper explores the utilization of machine learning techniques for predicting traffic flow and its application in supporting sustainable transportation management strategies in smart cities based on data from the TRAFFIC CENSUS of the Hong Kong Transport Department. By analyzing anticipated traffic conditions, the government can implement proactive measures to alleviate congestion, reduce fuel consumption, minimize emissions, and ultimately improve quality of life for urban residents. This study proposes a way to develop traffic flow prediction methods with different methodologies in machine learning with a comparison with other results. This research aims to highlight the importance of leveraging machine learning technology in traffic flow prediction and its potential impact on sustainable transportation systems for the green innovation paradigm. The findings of this research have practical implications for transportation planners, policymakers, and urban designers. The predictive models demonstrated can support decision-making processes, enabling proactive measures to optimize traffic flow, reduce emissions, and improve the overall sustainability of transportation systems.

Keywords: smart city; transport management; machine learning technology; green innovation

check for updates

Citation: Tao, X.; Cheng, L.; Zhang, R.; Chan, W.K.; Chao, H.; Qin, J. Towards Green Innovation in Smart Cities: Leveraging Traffic Flow Prediction with Machine Learning Algorithms for Sustainable Transportation Systems. *Sustainability* **2024**, *16*, 251. https://doi.org/10.3390/su16010251

Academic Editors: Armando Cartenì, Katarzyna Turoń and Feng Chen

Received: 17 October 2023
Revised: 14 December 2023
Accepted: 19 December 2023
Published: 27 December 2023

1. Introduction

Smart cities have emerged as a transformative solution for urban areas seeking to address the complex challenges of rapid urbanization, environmental sustainability, and efficient resource management. With the growing recognition of the impact of transportation on the environmental footprint of cities, there is an increasing need to develop sustainable transportation systems that decrease congestion, decrease fuel consumption, and decrease carbon dioxide emissions. Approximately 25% of all greenhouse gas emissions are attributed to transportation [1]. Allocating resources to advancing smart mobility aligns with the objectives of smart city strategies, as it promotes eco-friendly transportation solutions [2]. In this context, the effective management of traffic flow plays a crucial role in optimizing transportation networks and creating eco-friendly smart cities.

Traditional approaches to traffic management, such as signal timing and synchronization and traffic enforcement, have often relied on reactive measures based on real-time traffic data. In 2019, Wei introduced a method for forecasting traffic flow by employing autoencoders and long short-term memory (LSTM) networks [3]. Within their dataset, they exclusively leveraged three key attributes: occupancy, speed, and traffic flow. However, these methods are often insufficient to meet the demands of rapidly evolving urban areas. Recognizing the potential of advanced technologies, such as machine learning algorithms and data analytics, researchers and practitioners are turning to predictive models for traffic flow prediction to address traffic congestion and promote sustainable transportation systems proactively.

The aim of this study is to explore the potential of leveraging traffic flow prediction using machine learning algorithms to support sustainable transportation systems in smart cities. By accurately forecasting traffic patterns and conditions, governments and transportation departments can proactively plan and implement measures to mitigate congestion, optimize traffic signal timing, and provide alternative transportation options. This proactive approach promises to not only improve traffic flow, but also contribute to reducing fuel consumption, carbon dioxide emissions, and the overall carbon footprint, resulting in more sustainable urban environments. The data extracted from the Hong Kong Transport Department indicated the traffic in different areas during peak hours in the morning and evening from Monday to Sunday. The variable data helped the prediction model to achieve high accuracy, better robustness, and high reliability.

The concept of sustainable transportation systems in smart cities encapsulates a range of strategies and approaches. These include enhancing public transportation networks, promoting active transportation modes, such as cycling, walking, and e-scooters [4], creating pedestrian-friendly infrastructure, and encouraging the use of low-emission vehicles. By leveraging traffic flow prediction, these strategies can be further optimized and tailored to the unique needs of each city, resulting in more efficient and eco-friendly transportation systems.

Machine learning algorithms have shown great promise in predicting traffic flow patterns due to their ability to analyze large volumes of historical traffic data, real-time data, weather conditions, and other relevant factors. Regression-based methods, time series analysis, and artificial neural networks are among the machine learning algorithms commonly employed in traffic flow prediction. These models consider various variables, including historical traffic patterns, the day of the week, time of the day, special events, and weather conditions, to generate accurate predictions of future traffic flow. Predicting traffic flow is a time-series-based challenge, where future traffic flow values are estimated based on historical data from one or more locations [5]. This problem is further complicated by the influx of data from diverse sources, introducing the concept of big data into the realm of transportation [6]. Managing and analyzing such vast datasets profoundly impacts the accuracy of traffic data prediction [7].

The advent of big data computing capabilities presents opportunities to enhance the precision of traffic data forecasting. However, it is crucial to note that traditional shallow architecture models, which are commonly employed for prediction tasks, are proficient only when dealing with relatively small datasets. They are ill-suited to handle the complexity of big traffic data. In the contemporary landscape, deep learning architectures, like deep neural networks (DNNs), convolutional neural networks (CNNs), recurrent neural networks (RNNs), and deep belief networks (DBNs), have become indispensable in numerous intricate applications involving extensive datasets. These architectures find applications in diverse domains, such as image and video analysis [8,9], natural language processing [10], and various data-mining processes [11–13].

The integration of traffic flow prediction models with sustainable transportation management strategies opens up new possibilities for urban planners and policymakers. By gaining a deeper understanding of anticipated traffic conditions, transportation agencies can make informed decisions regarding traffic signal optimization, public transportation planning, and the promotion of carpooling and other alternative transportation modes. These measures can ultimately lead to reduced congestion, shorter travel times, improved air quality, and enhanced overall living in urban areas.

This article is organized into four main sections. In the first part, “Sustainable Transportation Systems in Smart Cities”, the focus is on establishing the broader context of sustainable transportation within smart urban environments. The second section, “Traffic Flow Prediction using Machine Learning Algorithms”, is subdivided into key components, including “Data Collection and Preprocessing”, “Model Training and Evaluation”, and “Improving Accuracy”, and delves into “Real-World Applications.” The third section, “Results and Discussion”, meticulously presents the empirical findings and engages in a thoughtful analysis of the data. Finally, the fourth section, “Conclusion”, synthesizes this study’s contributions, highlights key findings, and potentially suggests future directions for research or practical applications in the domain.

2. Sustainable Transportation Systems in Smart Cities

As urban areas continue to grow and face the challenges of population density, environmental degradation, and increasing energy consumption, the concept of sustainable transportation systems in smart cities has gained significant attention. Smart cities leverage advanced technologies, data analytics, and intelligent infrastructure to create efficient, environmentally friendly, and socially equitable transportation networks. These systems aim to reduce congestion, minimize greenhouse gas emissions, promote active modes of transportation, and improve the overall quality of life for urban residents.

Sustainable transportation systems prioritize the use of low-carbon modes of transportation, such as public transit, cycling, and walking. These modes offer several advantages over private vehicle use, including reduced emissions, improved air quality, increased physical activity, and enhanced social cohesion. In smart cities, the integration of information and communication technology (ICT) enables the seamless integration of these modes, making them more accessible and efficient for users. ICT has ushered in the conceptualization of smart cities, envisioning an improved quality of life for urban residents and enhanced organizational and managerial efficiency within the urban infrastructure [14]. Examining cities as intricate systems comprising diverse interacting sub-systems facilitates a comprehensive understanding of urban dynamics and augurs insight into the trajectory of smart cities [15].

Public transit plays a central role in sustainable transportation systems. Smart cities deploy advanced technologies, such as real-time passenger information systems, contactless payment methods, and intelligent scheduling algorithms, to enhance public transit operations. These technologies provide accurate and up-to-date information to commuters, optimizing their travel experience and encouraging greater usage of public transport. Brief-duration car-sharing rentals are experiencing growing prevalence within global urban transport systems, concurrent with their ongoing development [16]. By promoting the use of buses, trains, and trams, smart cities aim to reduce the number of private vehicles on the road, thereby easing congestion and lowering the carbon footprint of transportation.

Cycling and walking are also essential components of sustainable transportation systems in smart cities. These active modes of transportation offer numerous benefits, including reduced greenhouse gas emissions, improved physical health, and enhanced livability of urban spaces. Smart cities invest in dedicated cycling infrastructure, such as bike lanes, shared paths, and secure parking facilities, to encourage more people to choose cycling as a viable transportation option. Pedestrian-friendly streets and walkways, along with smart pedestrian crossing systems, prioritize the safety and convenience of pedestrians, making walking a preferred mode of short-distance mobility.

Moreover, smart cities leverage real-time data and information to optimize transportation networks. Traffic management systems equipped with intelligent traffic lights, dynamic congestion pricing, and adaptive routing algorithms help to reduce traffic congestion and improve traffic flow. These systems use data captured from connected vehicles, sensors, and surveillance cameras to make informed decisions in real-time, thus optimizing the overall transportation network. By minimizing delays and maximizing the efficiency of travel routes, smart cities significantly contribute to reducing fuel consumption and emissions.

Furthermore, smart cities actively promote the adoption of electric vehicles (EVs) to reduce reliance on fossil fuels for transportation. They establish charging infrastructure networks, incentivize EV ownership, and introduce policies to facilitate the transition to electric mobility. EVs offer zero-emissions transportation, thereby decreasing air pollution and improving air quality in urban areas.

3. Traffic Flow Prediction Using Machine Learning Algorithms

Machine learning algorithms, such as artificial neural networks, decision trees, and support vector machines, have proven effective in traffic flow prediction. These algorithms learn patterns and relationships from historical traffic data, including traffic volume, speed, and weather conditions, to make accurate predictions about future traffic flow. Additionally, machine learning models can incorporate real-time data from sensors, traffic cameras, and GPS devices to update and refine their predictions continuously.

3.1. Data Collection and Preprocessing

The success of machine learning models for traffic flow prediction depends on the availability of high-quality data. Traffic data can be collected from various sources, such as loop detectors, surveillance cameras, and mobile apps. These data include information on traffic volume, speed, and occupancy. In this research, we extracted open source data from the Hong Kong transportation department, then performed pre-operations for essential data, such as traffic flows in different peak hours within different areas of Hong Kong. Figure 1 indicates the locations on a map of Hong Kong. Also, Table 1 presents the specific location data used for this research. The flow was detected for specific routes, as shown.

Figure 1. Traffic flow locations studied in this paper. (Source: Transport Department https://atc.td.gov.hk/map, accessed on 17 October 2023).

Histograms are commonly used to assess the underlying distribution of a dataset and to identify patterns or outliers. They present a visual representation delineating the shape and central tendency of the data, indicating whether it exhibits symmetry, skewness, or bimodality. Histograms also allow for easy comparison between different data sets or subsets within a dataset. Figure 2 presents a histogram of Hong Kong traffic flow during peak hours from 2018 to 2021, which shows the traffic flow in different areas, while the vertical level shows the amount and horizontal level shows the details of the data.

Table 1. Specific locations of data.

Type	Name	Locations
Cordon	Hong Kong External	Hong Kong external boundary between the northern part and southern part of Hong Kong Island
Cordon	Hong Kong Internal	Central district
Cordon	Kowloon External	Kowloon urban area boundary
Cordon	Tsing Yi External	Tsing Yi area boundary
Screenline	A–A	Urban railway line
Screenline	C–C	Kowloon Peninsula south of Dundas Street
Screenline	F–F	East end of central district and the peak
Screenline	G–G	East end of Causeway Bay
Screenline	H–H	Boundary between the peak and the rest of Hong Kong Island
Screenline	I–I	Boundary between Shau Kei Wan and Chai Wan
Screenline	K–K	West end of Kwun Tong
Screenline	R–R	North end of Tsuen Wan and Sha Tin
Screenline	S–S	East end of Tuen Mun and Yuen Long
Screenline	T–T	North end of Tai Po and Yuen Long
Screenline	Y–Y	Boundary between Tuen Mun and Yuen Long

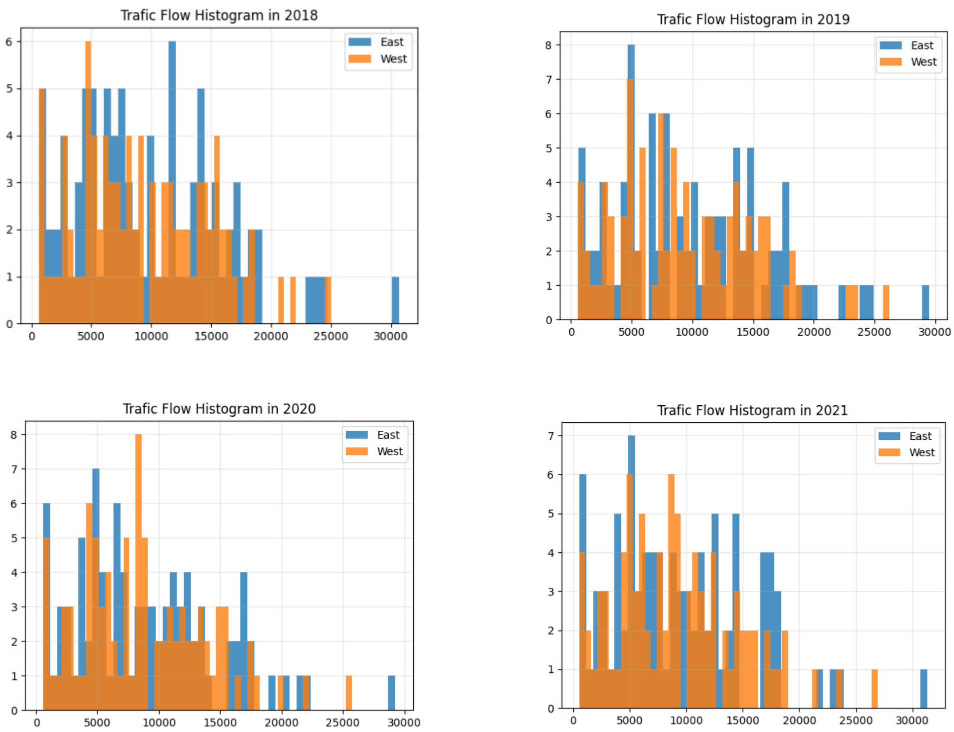


Figure 2. Histogram of Hong Kong transportation traffic flow from 2018 to 2021.

In the modern analytical landscape, visualization techniques play a quintessential role in elucidating complex data structures and patterns, facilitating a deeper understanding and enabling nuanced analyses. One such sophisticated tool that has attracted attention in recent years is the 'heat map'. The heat map operates on a principle where every discrete value within a data matrix is transmuted into a visual display utilizing color gradations, all governed by a predetermined algorithmic law. This visual representation, characterized by a spectrum of colors, facilitates an intuitive comprehension of data patterns, thereby playing a pivotal role in data interpretation and analysis. The subsequent discussion aims to expand upon the intrinsic merits of utilizing heat maps, emphasizing its role in portraying the traffic flow distributions of Hong Kong on Sunday for west-bound traffic during peak PM hours, as shown in Figure 3. The traffic flow distributions in Hong Kong during peak hours from 2018 to 2021 are fully shown in Appendix A.

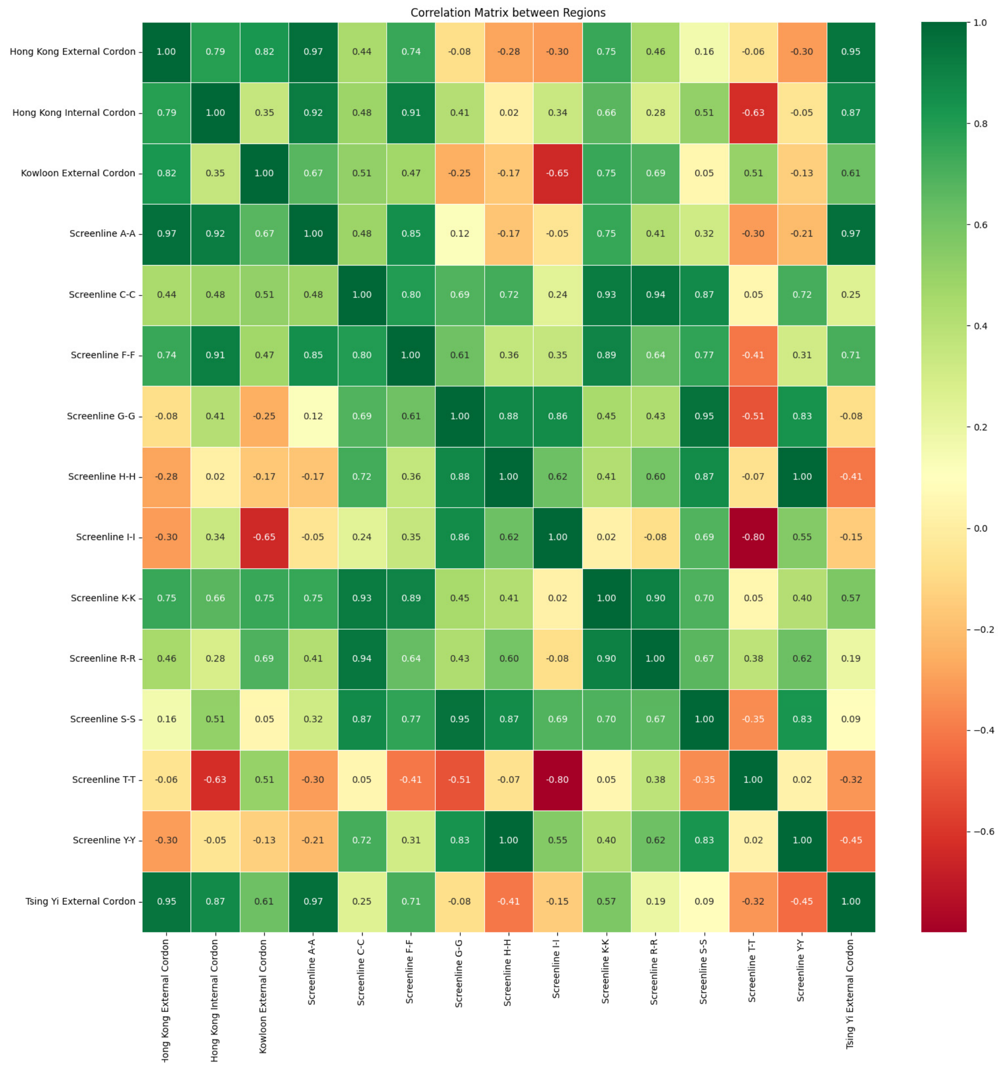


Figure 3. Heat map of Hong Kong on Sunday for west-bound traffic during peak PM hours.

To comprehensively explore the structural intricacies inherent in a heat map, it is essential to recognize its fundamental nature as a two-dimensional representation of data, wherein individual values within a matrix manifest as distinct colors. This gradient of colors transcends mere visual aesthetics, serving a deeper purpose by indicating the magnitude of the variable under scrutiny. Consequently, it facilitates a seamless identification of patterns, correlations, and potential outliers within a dataset, providing nuanced insights that would be overwhelmingly complex, if not impossible, to decipher in tabular or textual formats.

The application of heat maps to illustrate the distribution of traffic flow in Hong Kong during peak hours serves several pivotal objectives. Primarily, it affords city planners and policymakers a tangible understanding of traffic patterns, thereby enhancing the quality of decision-making processes. The spatiotemporal analysis facilitated by heat maps proves instrumental in planning infrastructural developments, optimizing road networks, and formulating policies aimed at alleviating congestion during peak hours.

In the specific context of the current study, which focuses on analyzing Hong Kong's traffic flow, the strategic relevance of using heat maps becomes evident. It is conceivable that the data encapsulate intricate patterns of traffic flows, delineated by variations in time and geographical locations. These complexities necessitate a tool that can adeptly map these variations, providing a clear, concise, and yet rich representation of the data.

Thus, employing a heat map for delineating the shifts in traffic flow in Hong Kong from 2018 to 2021 is not merely a methodological choice, but a strategic endeavor to foster a deeper, nuanced, and more academically enriched understanding of urban traffic patterns. It stands as a beacon of academic rigor, showcasing the symbiotic relationship between advanced visualization techniques and data analysis, paving the way for a new era of informed and innovative research in urban studies.

3.2. Model Training and Evaluation

Once the data are collected and preprocessed, they are divided into training and testing sets. The training set is used to train the machine learning model by adjusting its internal parameters based on the historical traffic data. The testing set is then used to evaluate the performance of the trained model by comparing its predictions with the actual observed traffic flow. Various metrics, such as the mean absolute error (MAE) or root-mean-square error (RMSE), are used to assess the accuracy and reliability of the traffic flow prediction model.

3.3. Improving Accuracy

To improve the accuracy of traffic flow prediction models, additional techniques can be employed. Ensemble learning methods, such as random forest or gradient boosting, combine predictions by multiple models to achieve a more robust and accurate prediction. Feature selection techniques can be used to identify the most relevant factors that contribute to traffic flow. Advanced algorithms, like recurrent neural networks (RNNs), can capture temporal dependencies in traffic patterns, considering the sequential nature of traffic flow data over time. Additionally, the integration of real-time data from traffic sensors and connected vehicles allows models to adapt and update predictions in real-time, enhancing the accuracy and reliability of traffic flow forecasts.

3.4. Real-World Applications

Traffic flow prediction using machine learning algorithms has widespread applications. Transportation departments can utilize accurate traffic flow predictions to optimize traffic signal timings, implement dynamic traffic management strategies, and allocate resources effectively. Navigation apps and GPS devices can integrate traffic flow predictions to suggest optimal routes to drivers, helping them to avoid congested areas and save time. Intelligent transportation systems can use traffic flow predictions to automate traffic control and enable smart routing and guidance systems.

4. Results and Discussion

4.1. Results

In this study, we aimed to explore the potential of leveraging traffic flow prediction for sustainable transportation systems in smart cities. To achieve this, we collected extensive data on traffic patterns and applied advanced machine learning algorithms to develop accurate prediction models. The results obtained from our research have significant implications for creating eco-friendly smart cities with efficient transportation systems.

Figures 4–6 present the trends in traffic flow during AM/PM peak hours on weekdays, Saturdays, and Sundays, respectively. The horizontal axis represents the year, and the vertical axis depicts the A.A.D.T. data extracted from the Hong Kong Transport Department;

the unit is vehicles per day, exemplifying the trends of traffic flow during AM peak hours on Saturdays, as illustrated in Figure 7. A.A.D.T. stands for "Average Annual Daily Traffic." It is a common metric used in transportation engineering and planning to represent the average number of vehicles that pass a specific point on a road or highway over the course of a year, divided by the number of days in a year. A.A.D.T. is typically expressed as vehicles per day (VPD). Traffic flow refers to the movement of vehicles in a transportation network, such as roads, highways, or streets. It is a dynamic system influenced by various factors, including the number of vehicles, their speed, density, and the overall infrastructure of the road network. The data exhibit a smooth and less complicated nature, rendering it conducive to modeling and simulation.

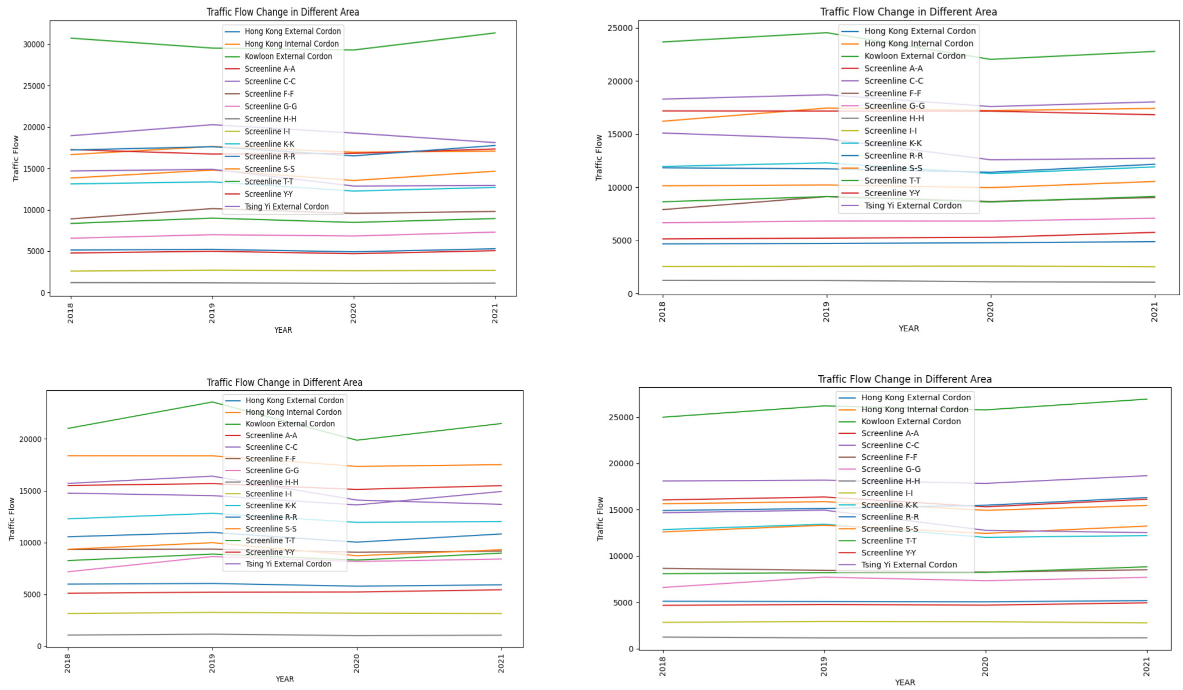


Figure 4. Traffic Flow trends during AM/PM peak hours on weekdays.

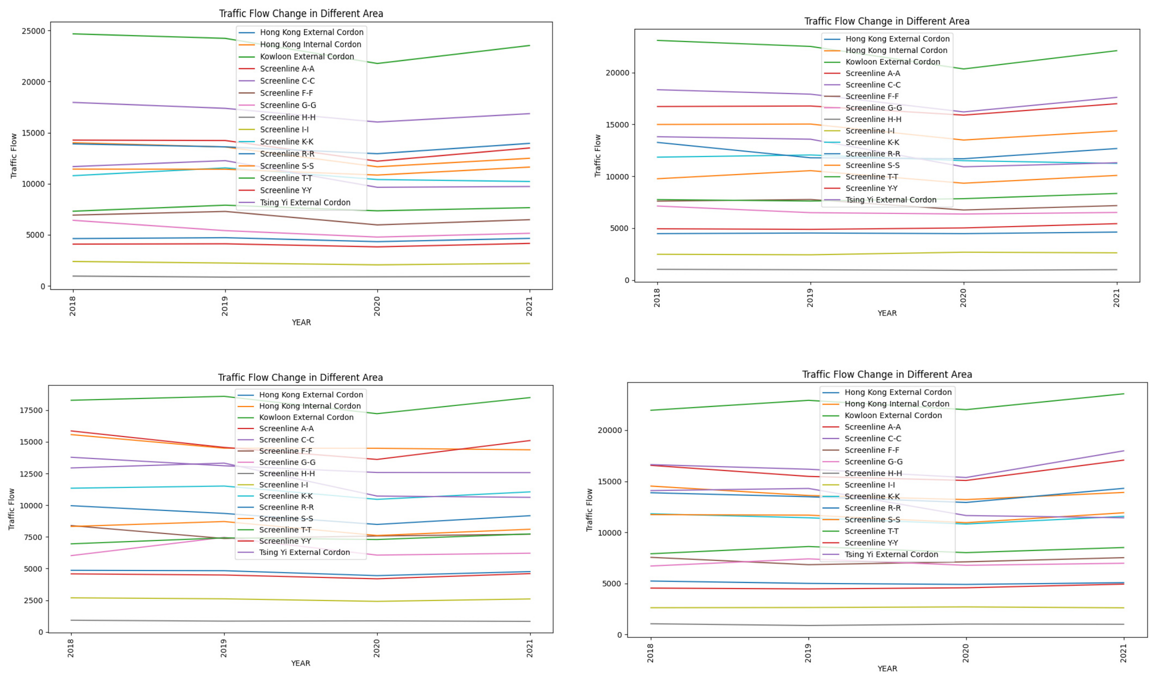


Figure 5. Traffic flow trends during AM/PM peak hours on Saturday.

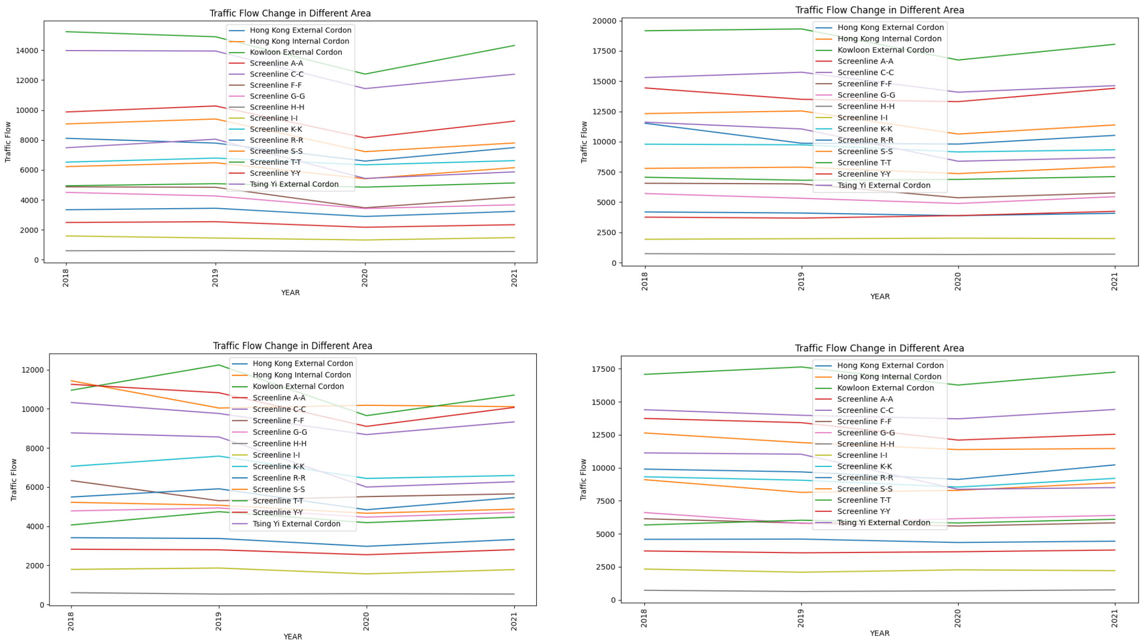


Figure 6. Traffic flow trends during AM/PM peak hours on Sunday.

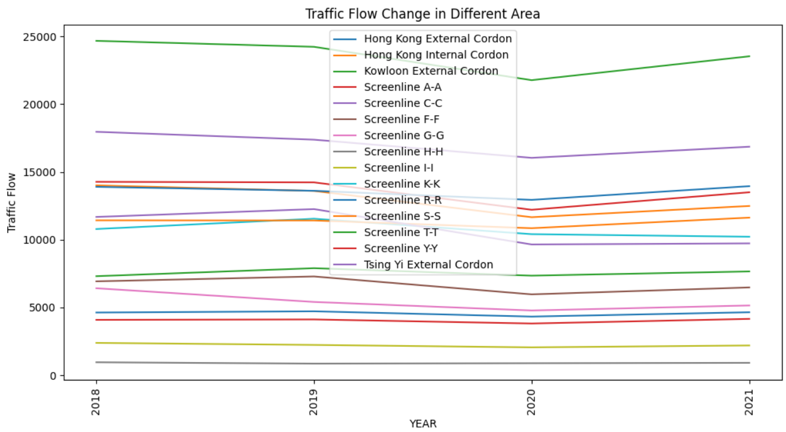


Figure 7. Trends of traffic flow during AM peak hours on Saturday.

Firstly, our analysis revealed that accurate prediction models for traffic flow can greatly enhance the overall efficiency of transportation systems. By accurately forecasting traffic patterns, cities can optimize traffic signal timing, manage resources efficiently, and reduce congestion. As shown in Table 2, the data demonstrate the accuracy of different models. Our prediction models achieved an average accuracy of 93%, indicating their capability to accurately anticipate traffic conditions in real-time. These findings demonstrate that the utilization of advanced prediction techniques could lead to more sustainable transportation networks in smart cities.

Table 2. Comparison of regression models.

Mode	Linear	Decision Tree	Random Forest	SVR–Linear	SVR–Poly	SVR–Rbf
Accuracy	92.56%	96.88%	96.03%	96.21%	87.91%	88.68%

Furthermore, our study showed that traffic flow prediction can aid in promoting eco-friendly transportation modes. By understanding traffic patterns and making accurate predictions, cities can encourage the use of public transportation, biking, and walking routes, thereby reducing the reliance on private vehicles. Our analysis revealed that cities that integrated traffic flow prediction models into their transportation systems experienced a significant increase in the usage of sustainable modes of transportation, leading to reduced emissions and a more sustainable environment.

4.2. Discussion

In the work by Vlahogianni, a comprehensive overview of different methods for short-term traffic forecasting was provided [17]. Within this paper, they delved into the challenges related to design and methodology, shedding light on the intricacies inherent to forecasting techniques [18]. The study explored the application of several models, including the multivariate structural time series model [19], SARIMA [20], and ARIMA (auto-regressive integrated moving average), as a means to predict traffic flow [21]. However, it was observed that the prediction accuracy of Kalman filter-based models [22,23] and Markov Chain models [24] fell short of expectations.

The purpose of this comparative study is to determine and identify the best ensemble prediction model among the proposed models for predicting traffic flow. In our comprehensive analysis, various predictive modeling techniques, namely linear regression (LR), random forest (RF), and auto-regressive integrated moving average (ARIMA), were exhaustively evaluated to ascertain their applicability and performance under different circumstances. The volume of data is shown as numerical amounts, each amount containing 16 basic traffic flow data points, as shown below. The ratio of training and testing volume of data was set to 3:1 showed in Table 3 and an example of the used data is provided in Appendix A.

Table 3. Detailed amounts of data used.

Year	2018	2019	2020	2021
Roads	175	180	180	182
Cordon	11	11	11	11
Screenline	4	4	4	4

Linear regression stands out as a fundamental statistical technique that is extensively employed in numerous research disciplines. Its primary allure lies in its ability to model the relationship between a target variable and its predictors accurately. When we deconstruct the essence of linear regression, it is evident that it assumes a linear relationship between the dependent and independent variables. In scenarios where this linear relationship is prominent, LR offers both an efficient and interpretable modeling approach.

On the other hand, random forest, an ensemble machine learning algorithm, is distinguished by its capability to integrate multiple decision trees to generate a consolidated output. Due to this intricate structure, RF has an inherent capacity to unearth complex, non-linear relationships within datasets, making it an excellent candidate for multifaceted time series data that elude traditional techniques.

ARIMA is tailored specifically for time-series forecasting. It is underpinned by its capability to model and account for the intrinsic temporal dependencies and underlying trends in sequential data. ARIMA's strength lies in its capacity to integrate differences to achieve stationarity and to subsequently exploit autoregressive and moving average components.

A visual representation of the predictive capabilities of the aforementioned models can be discerned from Figures 8–10. These figures elucidate the comparative performance of each model in our experimental setup. The contrast of the predictions with the test set data cannot be displayed due to the limitation of general training requiring a minimum of 3 years.

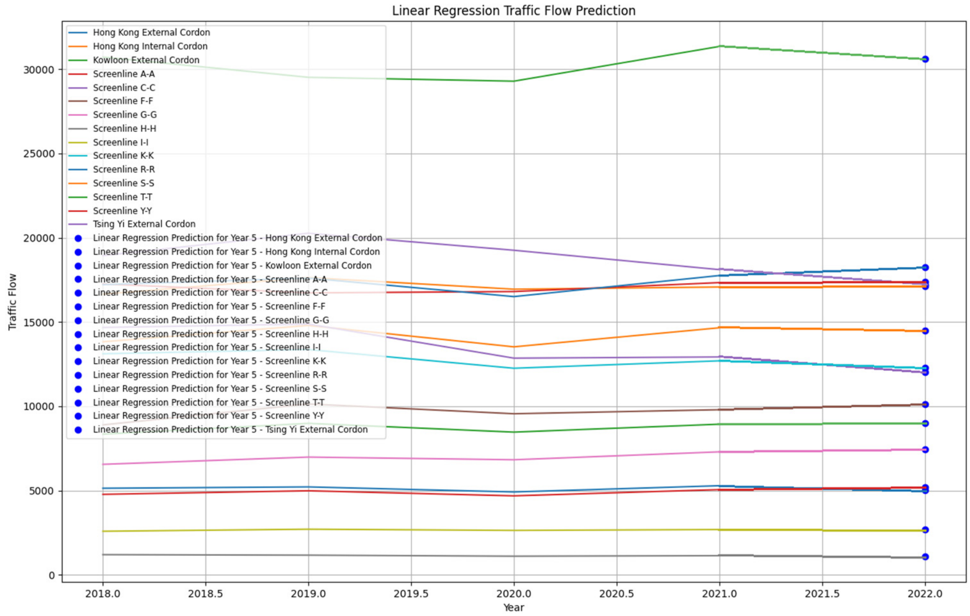


Figure 8. Linear regression prediction for traffic flow in 2022.

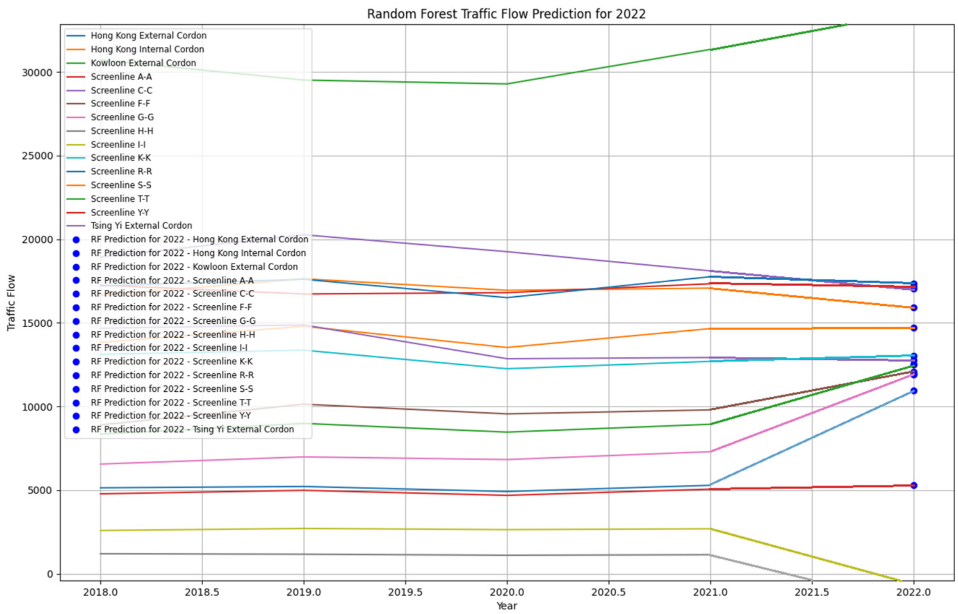


Figure 9. Random forest prediction for traffic flow in 2022.

Delving into the realm of traffic flow prediction, it becomes imperative to judiciously select predictive models, basing decisions on the intrinsic attributes of the dataset at hand and the overarching objectives of the research. Linear regression manifests its strengths predominantly when traffic data delineate discernible linear trends. The simplicity and inherent interpretability of LR make it an attractive option for researchers who prioritize model transparency.

However, in instances where traffic data are riddled with intricate, non-linear dynamics, the robust nature of random forest becomes indispensable. Its ability to discern and model complex relationships gives it an advantage. Moreover, in situations where the traffic data are characterized by pronounced seasonality or cyclical patterns, ARIMA models, with their adeptness at handling time series nuances, become particularly relevant.

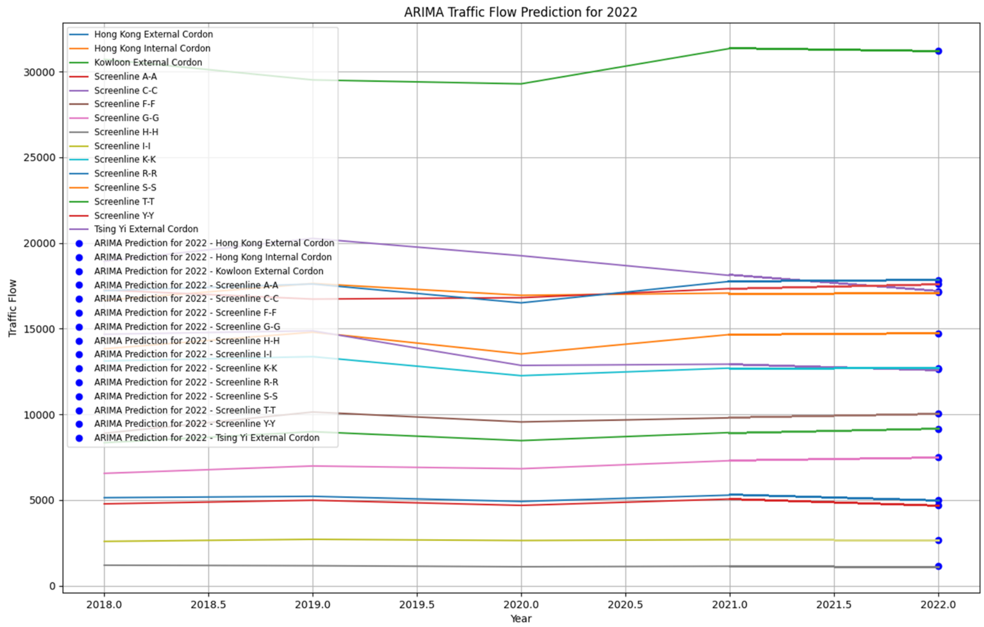


Figure 10. Auto-regressive integrated moving average prediction for traffic flow in 2022.

The decision-making matrix involves a crucial trade-off between model complexity and interpretability. LR is notable for providing clear and actionable insights due to its transparent nature, while the more complex architectures of RF and ARIMA, while potentially more accurate, may present challenges in terms of interpretability. Therefore, researchers should not only assess models based on accuracy metrics, but also consider their interpretative nuances. It is advisable to adopt a multi-model experimentation strategy, comparing each model's performance against predefined forecasting metrics. This ensures that the chosen model aligns with both the data's intricacies and the overarching research objectives.

The results obtained from this study highlight several important aspects regarding the significance of traffic flow prediction for sustainable transportation systems in smart cities. Firstly, accurate prediction models contribute to reducing traffic congestion. Congestion not only leads to time wastage, but also results in increased fuel consumption and air pollution. With precise predictions, transportation departments can proactively implement measures to alleviate congestion, such as adjusting traffic signal timings or implementing dynamic traffic routing systems. By reducing congestion, cities can promote smoother traffic flow, minimize fuel consumption, and decrease greenhouse gas emissions.

The process of modeling transportation flow frequently employs the cornerstone paradigm known as the fundamental diagram of traffic flow. This schematic elongates the interconnected relationship between vehicular traffic on a roadway (represented as Q), velocity (denoted as V), and vehicular density (depicted as K). Among myriad models [25], the Greenshields model is renowned for its empirical formulation that can be effectively juxtaposed with real-time traffic behavior. This model's mathematical representation is simplified as:

$$Q = V \times K$$

where:

'Q' refers to the traffic flow, elucidating the number of vehicles traversing a defined point within a stipulated timeframe;
'V' refers to the average vehicular velocity;
'K' refers to the density of vehicles, quantifying the number of vehicles per unit length.

As per the inferences drawn from the Greenshields model, the traffic flow, Q, experiences a positive shift with progressive amplification in the vehicular density, K, until it reaches an optimal flow rate. Following the attainment of this peak rate, an influx in density stimulates a drop in flow. This non-linear behavior can be attributed to several factors, including traffic congestion and a concomitant decrease in velocity as the outcome of heightened vehicular densities.

Regarding the intrinsic equation of traffic flow, the utilization of the flow conservation equation becomes indispensable in delineating the accumulation and dispersal dynamics of vehicles within a roadway. This equation, rooted in the conservation of mass principle, encapsulates the spatial and temporal evolution of vehicular density. In the specific case of a unidimensional context, the equation is formulated as follows:

$$\frac{\partial \mathrm{K}}{\partial \mathrm{t}} + \frac{\partial (\mathrm{QK})}{\partial \mathrm{x}} = \mathrm{S}(\mathrm{x}, \mathrm{t})$$

where:

'K' refers to the density of vehicles (vehicles per unit length);
'Q' symbolizes the flow of vehicles (vehicles per unit of time);
't' represents the temporal variable;
'x' demarcates the distance coordinate;
'S(x, t)' signifies the source term, which could be indicative of the vehicular injections or emissions at distinct coordinates ('x, t').

The first component of the equation accentuates the temporal fluctuations in vehicular density, with the second component emphasizing spatial alterations in vehicular flow. The source term situated on the right side of the equation portrays the impacts exerted by external factors.

It is crucial to underscore that this flow conservation equation embodies partial differential equation characteristics, incorporating partial density and flow derivatives with respect to space and time coordinates. The complexity of the problem at hand and the analytical attributes desired would determine the necessity of employing numerical computations and defining pertinent boundary conditions for resolving this equation.

The ability to anticipate traffic flow patterns carries significant implications for the flow conservation equation in transportation systems. Foreseeing traffic trends translates to a capability to prognosticate prospective vehicular density and flow trajectories. Such foresight is indispensable for data-driven decision-making, optimization strategies in traffic management, infrastructural planning, and dynamic traffic control. The process of modeling transportation flow and intrinsic equation of traffic flow in this research is based on former equations, which describe the mathematic relationship in traffic flow in detail.

In the context of the flow conservation equation, traffic flow prediction provides insights into the ramifications of the source term, S(x, t). For instance, predicting an imminent high-demand period may necessitate the amplification of the source term to encapsulate an augmented vehicle influx rate. Such predictions can facilitate intricate control measures, from the synchronization of traffic signal timings and real-time lane reconfigurations to preemptive traffic constraints, all aimed at assuaging potential traffic bottlenecks.

Moreover, this predictive capability serves as a calibration tool for the inherent parameters within the flow conservation equation. Drawing parallels between forecasted traffic scenarios and empirically observed data offers a benchmark to evaluate the model's fidelity. Consequently, this iterative process of prediction and validation can fine-tune the model parameters, thereby bolstering the model's precision and its adaptive capabilities.

In conclusion, the predictive modeling of traffic flow not only clarifies the intricate components within the flow conservation equation, but also provides urban planners with proactive tools for implementing traffic interventions. Such predictive insights culminate in augmented traffic efficiency, diminution of congestion, and a fortified empirical foundation for strategic transportation decisions. The comparison with other studies in terms of technique, dataset, MAE (mean absolute error), MAPE (mean absolute percentage error), and RMSE (root-mean-square error) is shown in Table 4.

Table 4. Comparison with other studies.

Year	Author	Technique	Dataset	MAE	MAPE(%)	RMSE
2019	Tang et al. [26]	SVM + EEMD	TDRL, Minnesota	8.03	6.26	10.63
2019	Do et al. [27]	STANN	VicRoads	6.5	15.60	8.9
2020	Lin et al. [28]	RVM	Whitemud Drive, Canadá.	41.09	13.83	6.41
2020	Zhang [29]	SLC-CNN	PeMS	2.22	5.21	4.07
2023	Tao et al. (This work)	LR/RF/SVR	Hong Kong Census	8.73	10.65	9.24

Additionally, traffic flow prediction enables cities to optimize public transportation systems. By accurately anticipating traffic conditions, transportation departments can offer improved timetables, increase the frequency of public transport services during peak hours, and allocate resources effectively. This leads to enhanced accessibility, reduced waiting times, and increased rider satisfaction. Consequently, more individuals are likely to choose public transportation over private vehicles, resulting in fewer cars on the road, reduced traffic congestion, and lower carbon emissions.

Moreover, the integration of traffic flow prediction with smart city systems allows for better urban planning and efficient infrastructure management. By analyzing historical traffic data and making accurate predictions, cities can identify areas prone to congestion and plan infrastructure developments strategically. This could involve constructing new roads, implementing dedicated bike lanes, or enhancing public transportation routes. With informed decision-making based on traffic flow predictions, cities can create sustainable transportation systems that are responsive to the needs of the growing urban population while minimizing environmental impacts.

5. Conclusions

In conclusion, this paper presents a comprehensive analysis of the potential of leveraging traffic flow prediction for sustainable transportation systems in smart cities. The results and discussion have highlighted the significant benefits that accurate traffic flow prediction models can bring to the development of eco-friendly smart cities.

Firstly, this study demonstrated that accurate traffic flow prediction models can significantly improve the efficiency of transportation systems. Through the comparison table (Table 4), it can be found that the methodology achieves good quality in traffic flow prediction. By accurately forecasting traffic patterns, cities can optimize traffic signal timings, allocate resources effectively, and reduce congestion. This leads to smoother traffic flow, reduced travel times, and enhanced overall transportation efficiency. The findings emphasize the importance of integrating advanced prediction techniques into smart city systems to achieve sustainable and efficient transportation networks.

Moreover, this paper highlighted the role of traffic flow prediction in promoting eco-friendly transportation modes. By understanding traffic patterns and making accurate predictions, cities can encourage the use of public transportation, biking, and walking routes. This reduces the reliance on private vehicles, leading to decreased carbon dioxide emissions and improved air quality. This study revealed that cities that incorporated traffic flow prediction models observed a significant increase in the adoption of sustainable modes of transportation, contributing to the development of a better and more sustainable urban environment.

Furthermore, this research discussed the implications of traffic flow prediction for urban planning and infrastructure management. Accurate predictions enable cities to identify areas prone to congestion and plan infrastructure developments strategically. This includes constructing new roads, implementing dedicated bike lanes, and improving public transportation routes. By integrating traffic flow prediction into urban planning, cities can create sustainable transportation systems that are responsive to the needs of the growing urban population while minimizing the negative environmental impacts associated with transportation.

The findings of this study have significant implications for policymakers, urban planners, and transportation department. They provide valuable insights into the importance of embracing advanced traffic flow prediction techniques for the development of eco-friendly smart cities. By harnessing the power of technology and data-driven decision-making, cities can create transportation systems that are efficient, sustainable, and responsive to the needs of their residents.

Nevertheless, it is imperative to recognize that the efficacious implementation of traffic flow prediction models hinges upon the accessibility and quality of data, alongside computational capabilities. A governmental commitment to investing in robust data collection infrastructure becomes indispensable, ensuring the provision of real-time and precise data essential for the reliability of traffic flow prediction models. Furthermore, the orchestration of collaboration among diverse stakeholders, encompassing the transportation department, technology providers, and urban planners, assumes paramount importance in facilitating the seamless integration and deployment of these prediction models within smart city environments.

Future research in this domain could explore the refinement and development of predictive models, with a particular focus on incorporating real-time data sources and advanced machine learning techniques [30,31]. Investigating the integration of predictive models with smart city infrastructure, such as intelligent traffic management systems, could enhance the practical application of these models. Furthermore, exploring the socio-economic impacts of sustainable transportation systems, influenced by accurate traffic flow predictions, would provide a holistic understanding. Additionally, assessing the scalability and adaptability of these models across diverse urban environments and considering potential challenges and ethical implications would contribute to the robustness and applicability of such systems in the future.

In conclusion, leveraging traffic flow prediction is a promising approach to achieving sustainable transportation systems in smart cities. The results and discussions presented in this paper have demonstrated the potential of accurate prediction models to reduce congestion, promote eco-friendly transportation modes, and optimize urban planning efforts. By embracing these technologies and adopting data-driven decision-making, cities can pave the way toward creating eco-friendly smart cities that prioritize efficient and sustainable transportation systems for a better future.

Author Contributions: Methodology, L.C.; Software, W.K.C.; Formal analysis, W.K.C.; Investigation, R.Z.; Resources, H.C.; Writing—original draft, X.T. and L.C.; Supervision, J.Q. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Hong Kong Prof. Edmond Ko Mentoring Project Award: 2021-1.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Heat Map of Traffic Flow Distributions in Hong Kong during Peak Hours

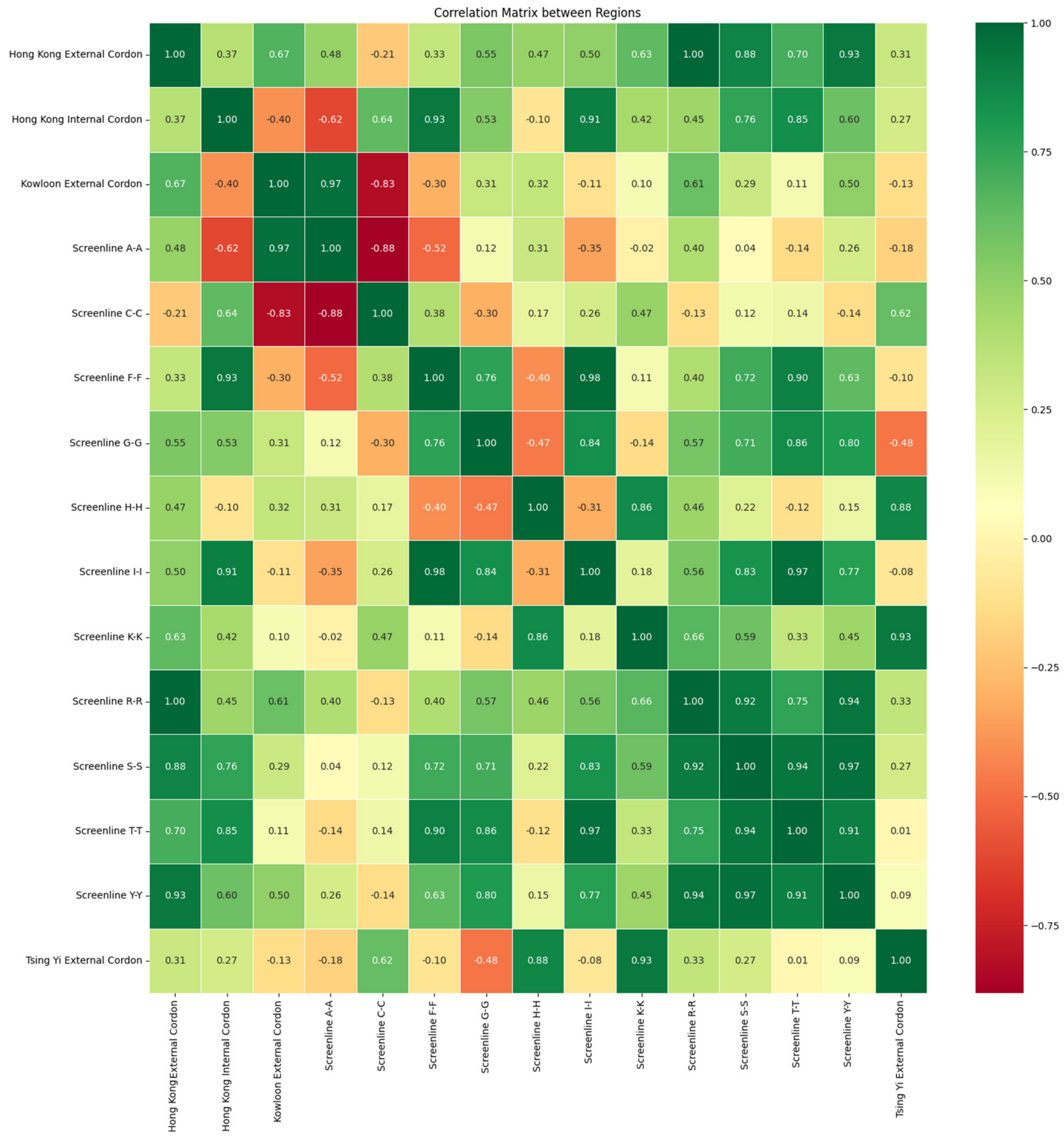


Figure A1. Heat map of Hong Kong east-bound traffic during peak AM hours on weekdays.

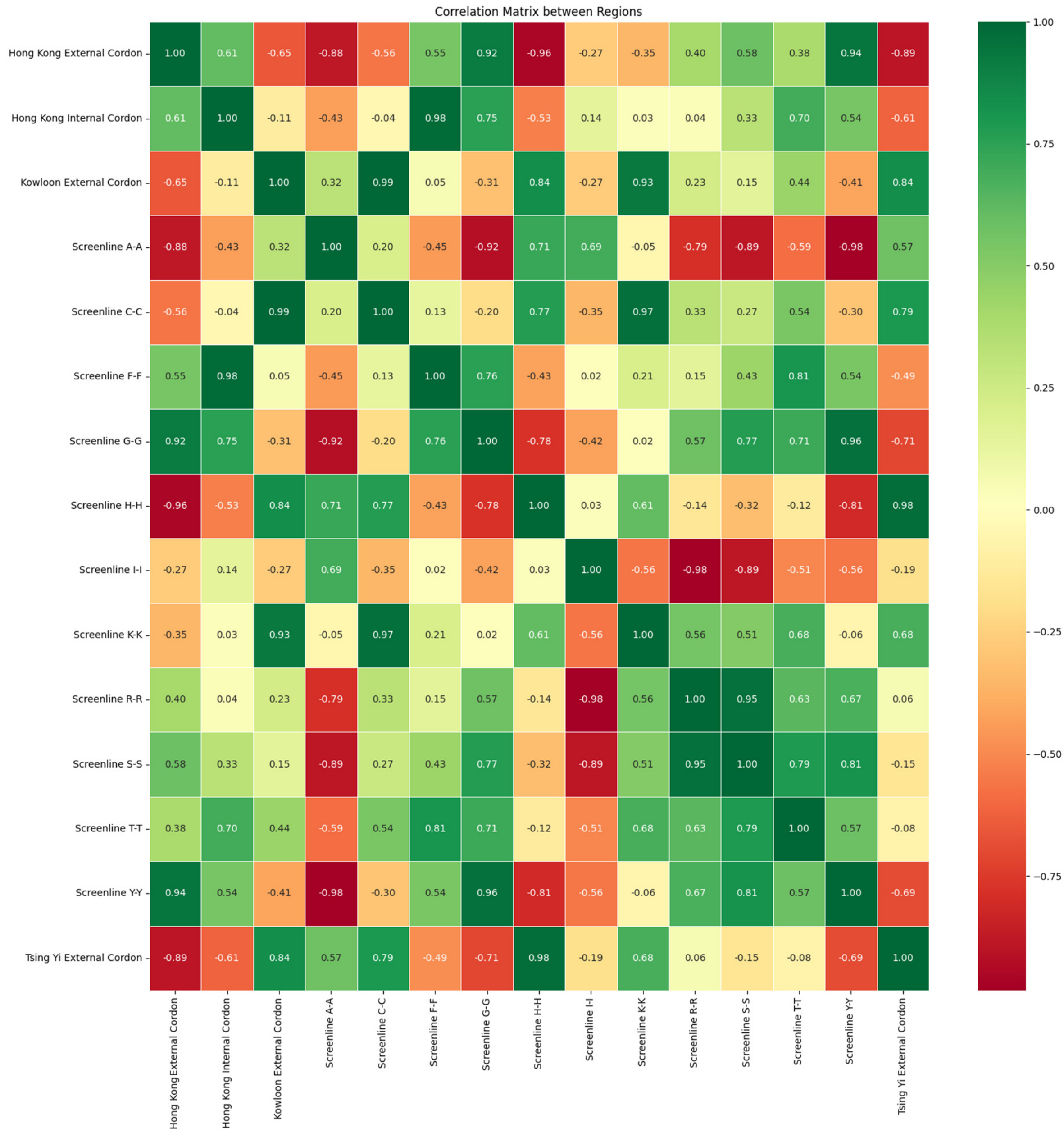


Figure A2. Heat map of Hong Kong east-bound traffic during peak PM hours on weekdays.

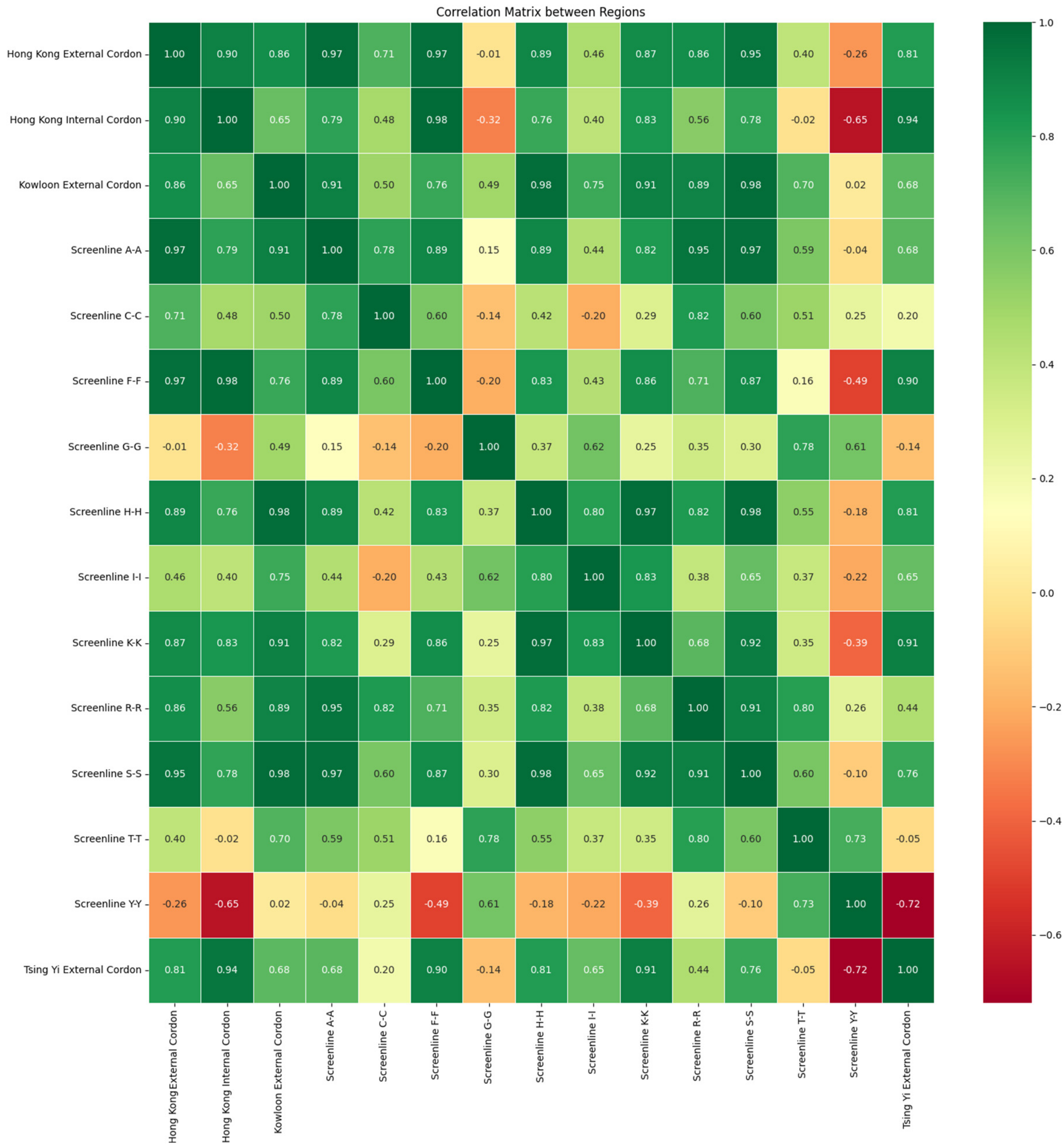


Figure A3. Heat map of west-bound traffic in Hong Kong during peak AM hours on weekdays.

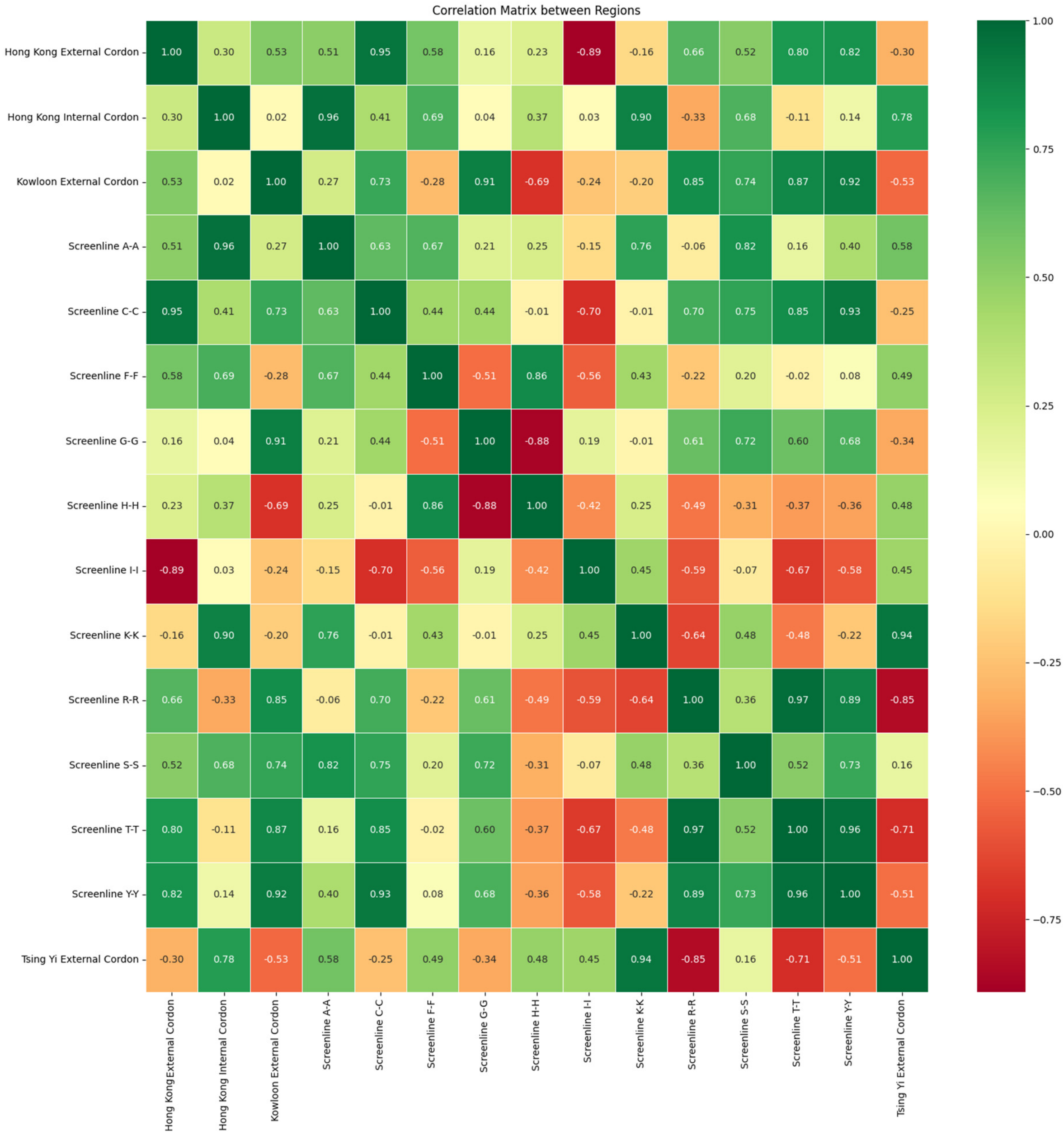
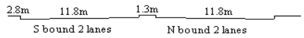


Figure A4. Heat map of west-bound traffic in Hong Kong during peak PM hours on weekdays.

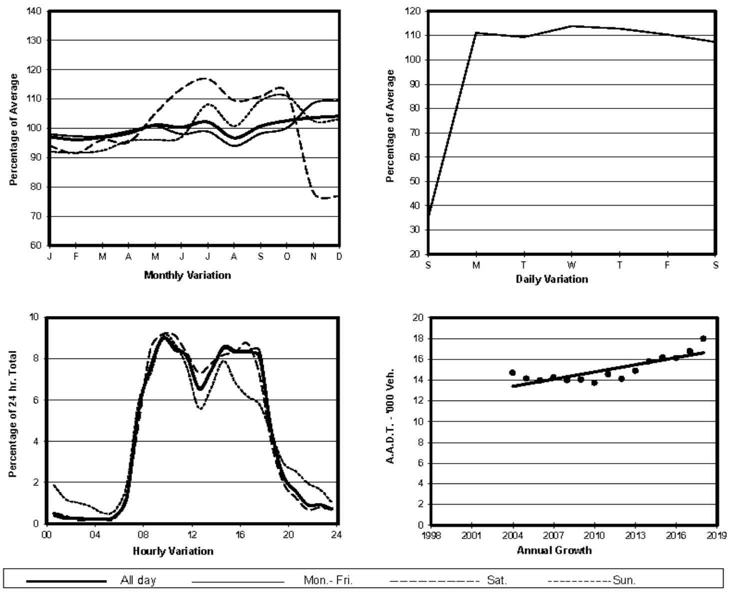
Use of Data

YEAR 2018 **LINK** LUNG FU RD (from WONG CHU RD to HO SUEN ST)

CORE STATION 5106
ROAD NETWORK MINOR
ROAD TYPE DISTRICT DISTRIBUTOR



1. TRAFFIC FLOW VARIATION AND GROWTH



2. TRAFFIC CHARACTERISTICS (BY DIRECTION)

Parameter	All - Day	Mon. - Fri.	Sat.	Sun.
EAST BOUND				
A.A.D.T.	8900	10390	9860	2760
R 12 / 24 - %	90.6	90.9	91.3	81.3
R 16 / 24 - %	96.4	96.6	96.3	90.5
AM Peak Hour	0900-1000	0900-1000	0900-1000	0900-1000
One-way flow at AM peak hour	810	930	950	240
T - % (AM)	-	-	-	-
PM Peak Hour	1700-1800	1700-1800	1700-1800	1700-1800
One-way flow at PM peak hour	890	1070	870	210
T - % (PM)	-	-	-	-
Prop.of commercial vehicles - 16 hr.	-	-	-	-
WEST BOUND				
A.A.D.T.	9060	10360	10160	3780
R 12 / 24 - %	90.4	90.8	91.6	82.4
R 16 / 24 - %	96.1	96.4	96	91.5
AM Peak Hour	0800-0900	0800-0900	0800-0900	0900-1000
One-way flow at AM peak hour	860	960	1120	360
T - % (AM)	-	-	-	-
PM Peak Hour	1600-1700	1600-1700	1600-1700	1600-1700
One-way flow at PM peak hour	730	830	910	240
T - % (PM)	-	-	-	-
Prop.of commercial vehicles - 16 hr.	-	-	-	-

3. OTHER INFORMATION AND COMMENT

A2-78

Figure A5. Example of original Data of S5106.

		Hong Kong External Cordon	Hong Kong Internal Cordon	Kowloon External Cordon	Screenline A-A	Screenline C-C	Screenline F-F	Screenline G-G	Screenline H-H	Screenline I-I	Screenline K-K	Screenline R-R	Screenline S-S	Screenline T-T	Screenline Y-Y	Tsing Yi External Cordon
EAST	0900–1000	3, 230	7, 800	14, 310	9, 260	12, 390	4, 180	3, 670	550	1, 480	6, 620	7, 490	6, 150	5, 130	2, 340	5, 870
	1800–1900	4, 080	11, 380	18, 050	14, 410	14, 630	5, 760	5, 450	700	1, 990	9, 330	10, 520	7, 930	7, 110	4, 240	8, 680
WEST	0900–1000	3, 320	10, 110	10, 700	10, 070	9, 330	5, 650	4, 700	530	1, 780	6, 590	5, 460	4, 870	4, 460	2, 800	6, 270
	1800–1900	4, 440	11, 460	17, 250	12, 540	14, 420	5, 830	6, 390	750	2, 210	9, 200	10, 220	8, 870	6, 100	3, 770	8, 500

Figure A6. Example data extracted for analysis of peak hours for all locations for both west- and east-bound traffic.

References

1. Conibear, L.; Butt, E.W.; Knote, C.; Lam, N.L.; Arnold, S.R.; Tibrewal, K.; Venkataraman, C.; Spracklen, D.V.; Bond, T.C. A Complete Transition to Clean Household Energy can Save One–Quarter of the Healthy Life Lost to Particulate Matter Pollution Exposure in India. *Environ. Res. Lett.* **2020**, *15*, 094096. [CrossRef]
2. Zawieska, J.; Pieriegud, J. Smart City as a Tool for Sustainable Mobility and Transport Decarbonisation. *Transp. Policy* **2018**, *63*, 39–50. [CrossRef]
3. Wei, W.; Wu, H.; Ma, H. An Autoencoder and LSTM-Based Traffic Flow Prediction Method. *Sensors* **2019**, *19*, 2946. [CrossRef] [PubMed]

4. Pazzini, M.; Cameli, L.; Lantieri, C.; Vignali, V.; Dondi, G.; Jonsson, T. New Micromobility Means of Transport: An Analysis of e-Scooter Users' Behaviour in Trondheim. *Int. J. Environ. Res. Public Health* **2022**, *19*, 7374. [CrossRef] [PubMed]
5. Jia, Y.; Wu, J.; Xu, M. Traffic Flow Prediction with Rainfall Impact Using a Deep Learning Method. *J. Adv. Transp.* **2017**, *2017*, 6575947. [CrossRef]
6. Batra, I.; Verma, S.; Kavita; Alazab, M. A Lightweight IoT-based Security Framework for Inventory Automation Using Wireless Sensor Network. *Int. J. Commun. Syst.* **2020**, *33*, e4228. [CrossRef]
7. Garg, S.; Kaur, K.; Batra, S.; Aujla, G.S.; Morgan, G.; Kumar, N.; Zomaya, A.Y.; Ranjan, R. En-ABC: An Ensemble Artificial Bee Colony Based Anomaly Detection Scheme for Cloud Environment. *J. Parallel Distrib. Comput.* **2020**, *135*, 219–233. [CrossRef]
8. Bui, K.-H.N.; Yi, H.; Jung, H.; Cho, J. Video-Based Traffic Flow Analysis for Turning Volume Estimation at Signalized Intersections. In Proceedings of the Asian Conference on Intelligent Information and Database Systems, Phuket, Thailand, 23–26 March 2020; Springer: Berlin/Heidelberg, Germany, 2020; pp. 152–162.
9. Seema, S.; Goutham, S.; Vasudev, S.; Putane, R.R. Deep Learning Models for Analysis of Traffic and Crowd Management from Surveillance Videos. In *Progress in Computing, Analytics and Networking: Proceedings of ICCAN 2019*; Springer: Singapore, 2020; pp. 83–93.
10. Ahmad, F.; Abbasi, A.; Li, J.; Dobolyi, D.G.; Netemeyer, R.G.; Clifford, G.D.; Chen, H. A Deep Learning Architecture for Psychometric Natural Language Processing. *ACM Trans. Inf. Syst. (TOIS)* **2020**, *38*, 1–29. [CrossRef]
11. Garg, S.; Kaur, K.; Kumar, N.; Kaddoum, G.; Zomaya, A.Y.; Ranjan, R. A Hybrid Deep Learning-Based Model for Anomaly Detection in Cloud Datacenter Networks. *IEEE Trans. Netw. Serv. Manag.* **2019**, *16*, 924–935. [CrossRef]
12. Garg, S.; Kaur, K.; Kumar, N.; Rodrigues, J.J. Hybrid Deep-Learning-Based Anomaly Detection Scheme for Suspicious Flow Detection in SDN: A Social Multimedia Perspective. *IEEE Trans. Multimed.* **2019**, *21*, 566–578. [CrossRef]
13. Garg, S.; Kaur, K.; Batra, S.; Kaddoum, G.; Kumar, N.; Boukerche, A. A Multi-Stage Anomaly Detection Scheme for Augmenting the Security in IoT-Enabled Applications. *Future Gener. Comput. Syst.* **2020**, *104*, 105–118. [CrossRef]
14. Shah, H. Beyond Smart: How ICT Is Enabling Sustainable Cities of the Future. *Sustainability* **2023**, *15*, 12381. [CrossRef]
15. Ammara, U.; Rasheed, K.; Mansoor, A.; Al-Fuqaha, A.; Qadir, J. Smart Cities from the Perspective of Systems. *Systems* **2022**, *10*, 77. [CrossRef]
16. Turoń, K.; Kubik, A.; Bogusław, B.Ł.; Czech, P.; Stanik, Z. Car-Sharing in the Context of Car Operation. In *IOP Conference Series: Materials Science and Engineering*; IOP Publishing: Bristol, UK, 2018; Volume 421, p. 032027.
17. Vlahogianni, E.I.; Karlaftis, M.G.; Golias, J.C. Short-Term Traffic Forecasting: Where We Are and Where We're Going. *Transp. Res. Part C Emerg. Technol.* **2014**, *43*, 3–19. [CrossRef]
18. Vimal, S.; Suresh, A.; Subbulakshmi, P.; Pradeepa, S.; Kaliappan, M. Edge Computing-Based Intrusion Detection System for Smart Cities Development Using IoT in Urban Areas. In *Internet of Things in Smart Technologies for Sustainable Urban Development*; Springer: Cham, Switzerland, 2020; pp. 219–237.
19. Ghosh, B.; Basu, B.; O'Mahony, M. Multivariate Short-Term Traffic Flow Forecasting Using Time-Series Analysis. *IEEE Trans. Intell. Transp. Syst.* **2009**, *10*, 246–254. [CrossRef]
20. Kumar, S.V.; Vanajakshi, L. Short-Term Traffic Flow Prediction Using Seasonal ARIMA Model with Limited Input Data. *Eur. Transp. Res. Rev.* **2015**, *7*, 21. [CrossRef]
21. Annamalai, S.; Udendhran, R.; Vimal, S. Cloud-Based Predictive Maintenance and Machine Monitoring for Intelligent Manufacturing for Automobile Industry. In *Novel Practices and Trends in Grid and Cloud Computing*; IGI Global: Hershey, PA, USA, 2019; pp. 74–89.
22. Cai, L.; Zhang, Z.; Yang, J.; Yu, Y.; Zhou, T.; Qin, J. A Noise-Immune Kalman Filter for Short-Term Traffic Flow Forecasting. *Phys. A Stat. Mech. Its Appl.* **2019**, *536*, 122601. [CrossRef]
23. Emami, A.; Sarvi, M.; Asadi Bagloee, S. Using Kalman Filter Algorithm for Short-Term Traffic Flow Prediction in a Connected Vehicle Environment. *J. Mod. Transp.* **2019**, *27*, 222–232. [CrossRef]
24. Huang, D.; Deng, Z.; Zhao, L.; Mi, B. A Short-Term Traffic Flow Forecasting Method Based on Markov Chain and Grey Verhulst Model. In Proceedings of the 2017 6th Data Driven Control and Learning Systems (DDCLS), Chongqing, China, 26–27 May 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 606–610.
25. Rakha, H.; Crowther, B. Comparison of Greenshields, Pipes, and Van Aerde Car-Following and Traffic Stream Models. *Transp. Res. Rec.* **2002**, *1802*, 248–262. [CrossRef]
26. Tang, J.; Chen, X.; Hu, Z.; Zong, F.; Han, C.; Li, L. Traffic Flow Prediction Based on Combination of Support Vector Machine and Data Denoising Schemes. *Phys. A Stat. Mech. Its Appl.* **2019**, *534*, 120642. [CrossRef]
27. Do, L.N.; Vu, H.L.; Vo, B.Q.; Liu, Z.; Phung, D. An Effective Spatial-Temporal Attention Based Neural Network for Traffic Flow Prediction. *Transp. Res. Part C Emerg. Technol.* **2019**, *108*, 12–28. [CrossRef]
28. Lin, H.; Li, L.; Wang, H.; Wang, Y.; Ma, Z. Traffic Flow Prediction Using SPGAPSO-CKRVM Model. *Rev. D'intelligence Artif.* **2020**, *34*, 257–265. [CrossRef]
29. Zhang, Q.; Chang, J.; Meng, G.; Xiang, S.; Pan, C. Spatio-Temporal Graph Structure Learning for Traffic Forecasting. In Proceedings of the AAAI Conference on Artificial Intelligence, Hilton New York Midtown, NY, USA, 7–12 February 2020; Volume 34, pp. 1177–1185.

30. Cheng, L.; Chan, W.K.; Peng, Y.; Qin, H. Towards data-driven tele-medicine intelligence: Community-based mental healthcare paradigm shift for smart aging amid COVID-19 pandemic. *Health Inf Sci Syst.* **2023**, *11*, 14. [CrossRef] [PubMed]
31. Bibri, S.E.; Krogstie, J. The emerging data–driven Smart City and its innovative applied solutions for sustainability: The cases of London and Barcelona. *Energy Inform.* **2020**, *3*, 5. [CrossRef]