International Journal of
Geo-Information

MDPI

Review

Emerging Technologies for Smart Cities' Transportation: Geo-Information, Data Analytics and Machine Learning Approaches

Kenneth Li-Minn Ang [1,*], Jasmine Kah Phooi Seng [2,3], Ericmoore Ngharamike [4] and Gerald K. Ijemaru [1]

[1] School of Science, Technology and Engineering, University of the Sunshine Coast, Petrie, QLD 4502, Australia; gerald.ijemaru@research.usc.edu.au
[2] School of AI and Advanced Computing, Xian Jiaotong Liverpool University, Suzhou 215123, China; Jasmine.Seng@xjtlu.edu.cn
[3] School of Computer Science, Queensland University of Technology, Brisbane, QLD 4000, Australia; kahphooi.seng@qut.edu.au
[4] Department of Computer Science, Federal University Oye-Ekiti, Oye Ekiti 371104, Ekiti, Nigeria; ericmoore.ngharamike@fuoye.edu.ng
* Correspondence: lang@usc.edu.au

Abstract: With the recent increase in urban drift, which has led to an unprecedented surge in urban population, the smart city (SC) transportation industry faces a myriad of challenges, including the development of efficient strategies to utilize available infrastructures and minimize traffic. There is, therefore, the need to devise efficient transportation strategies to tackle the issues affecting the SC transportation industry. This paper reviews the state-of-the-art for SC transportation techniques and approaches. The paper gives a comprehensive review and discussion with a focus on emerging technologies from several information and data-driven perspectives including (1) geoinformation approaches; (2) data analytics approaches; (3) machine learning approaches; (4) integrated deep learning approaches; (5) artificial intelligence (AI) approaches. The paper contains core discussions on the impacts of geo-information on SC transportation, data-driven transportation and big data technology, machine learning approaches for SC transportation, innovative artificial intelligence (AI) approaches for SC transportation, and recent trends revealed by using integrated deep learning towards SC transportation. This survey paper aimed to give useful insights to researchers regarding the roles that data-driven approaches can be utilized for in smart cities (SCs) and transportation. An objective of this paper was to acquaint researchers with the recent trends and emerging technologies for SC transportation applications, and to give useful insights to researchers on how these technologies can be exploited for SC transportation strategies. To the best of our knowledge, this is the first comprehensive review that examines the impacts of the various five driving technological forces—geoinformation, data-driven and big data technology, machine learning, integrated deep learning, and AI—in the context of SC transportation applications.

Keywords: geo-information; transportation; smart cities; machine learning; data analytics; big data; deep learning; artificial intelligence (AI)

check for updates

Citation: Ang, K.L.-M.; Seng, J.K.P.; Ngharamike, E.; Ijemaru, G.K. Emerging Technologies for Smart Cities' Transportation: Geo-Information, Data Analytics and Machine Learning Approaches. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 85. https://doi.org/10.3390/ijgi11020085

Academic Editor: Wolfgang Kainz

Received: 1 December 2021
Accepted: 15 January 2022
Published: 24 January 2022

1. Introduction

The initial concept of a SC has been acknowledged as a framework that builds upon the advancements in the ICT (information and communication technology) field to address urbanization challenges. People are increasingly engaging with SC platforms in multiple ways (e.g., mobile devices, connected cars, smart homes). However, the development of frameworks for SCs has not fully matured to be able to take advantage of new and emerging data-driven technologies. The advancement of new technologies in big data, AI, machine learning, deep learning, and internet of things (IoT) will further shape the framework of a SC and revolutionize the different sectors in SCs [1,2]. Geoinformation and communication technology (GeoICT) [3] is another emerging field which is increasingly being utilized to foster urban sustainability and SCs. GeoICT has significant importance for the implementation of ICTs, involving geographic information science and systems in SCs to support analysis and decision-making.

SCs involve various ICTs and advanced technologies, which can transform many socio-economic aspects of society including health, energy, education, and transportation, thus enabling smart technologies to create change in society. The SC transportation industry is bound to face myriads of technological challenges as a result of unprecedented urban migration. Hence, it is expedient to devise efficient strategies to utilize the available infrastructure and to minimize traffic. Smart transportation systems play an important role in urban areas to address issues such as traffic control and urban congestion. Smart transportation systems can provide services to improve road safety, reduce accidents, and give on-time information to drivers and users. An example of a SC deployment can be found in the proposal by Alphabet (Google) to build public WiFi kiosks on streets in New York, with the potential to exchange data with autonomous vehicles & other urban systems [4].

SCs utilize a variety of tools and techniques, including technologies that rely on Intelligent Transportation Systems (ITS), big data, and data analytics, as well as AI, machine learning, deep learning, IoT, and edge analytics. SC technology-focused research addresses several research areas in smart transportation and its applications, which are significant components of SC requiring intelligent instrumentations and interconnections. These applications in transportation include driver experience, autonomous vehicles, collaborative traffic control, and management and traffic flow prediction.

In recent years, big data analytics has been utilized in the design and planning of smart transportation, control systems, and communities. In smart transportation, data is obtained from multiple heterogeneous sources such as GPS data, transportation logistic data, video data, social media data, sensors, and systems data e.g., vehicle-sensing data (VSD), vehicular mobile service data, advanced driver-assistance data, connected cars data, etc. A generic architecture of utilizing and deploying big data analytics in smart transportation systems is shown in Figure 1. The architecture has three layers for data sensing and collection, data analytics, and smart transportation application.

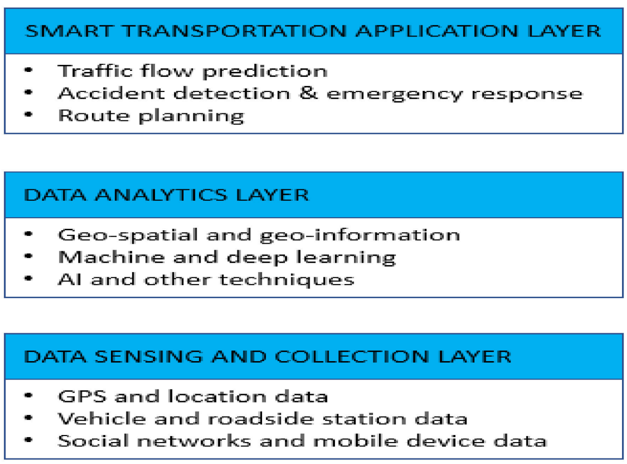

Figure 1. Generic architecture for big data analytics in smart transportation.

In the field of data science, machine learning is used for modelling and analytics to derive trends and patterns from data. In general, there are three categories for machine learning algorithms: (1) supervised learning algorithms; (2) unsupervised learning algorithms; (3) reinforcement learning algorithms. Neural networks (NN) are a popular supervised learning technique for both classification and regression [5]. Supervised learning algorithms require the use of labelled data for classification and regression. Unsupervised learning algorithms do not require labelled data and have been used for different smart transportation applications, such as traffic flow prediction [6,7], transportation travel route evaluation [8], parking spaces forecasting [9], bus arrival time prediction [10], etc.

A recent trend in modern machine learning algorithms is the emergence of deep learning (DL) models. Some DL models are the convolutional neural network (CNN), deep restricted Boltzmann machine, recurrent neural network, deep reinforcement learning models, stacked auto-encoders, etc. A survey of deep reinforcement learning for intelligent transportation can be found in [11]. Some applications include vehicle detection [12], traffic data imputation [13], and prediction of traffic flow density [14,15]. This paper focuses on innovative AI techniques for traffic modelling and prediction, traffic management and control, transportation and mobility, public transportation and other related applications for SCs.

The motivation behind this study was to investigate the impacts of the five driving, emerging technological forces—geoinformation, big data analytics, machine learning, integrated deep learning, and AI approaches—in the context of SC transportation strategies. The study was aimed to acquaint researchers with the recent trends and useful insights into the emerging technologies for SC transportation applications and using the state-of-the-art techniques to address the enormous challenges of the SC transportation industry. Although different methods have considered adopting each of the various techniques for the transportation industry, none have considered addressing the enormous challenges in the context of SC transportation strategies, using a combination of the various approaches. Due to the large volumes of real-time data being generated daily as a result of a rapid and unprecedented surge in urban migration, the existing conventional data processing tools are deficient to effectively realise the key targets of a SC transportation ecosystem. Consequently, this has brought enormous challenges for the SC transportation sector, including in traffic congestion, route planning issues, fleet management problems, parking request modelling problems, short-term forecasting problems, as well as the development of efficient strategies to utilize available infrastructures to minimize traffic/accidents and improve road safety. Thus, there is need to address these challenges by considering the integration of big data technology with other emerging technologies into the transportation sector and exploiting them for SC transportation applications. Hence, this paper considered the need to devise a state-of-the-art integrated approach to tackle the enormous challenges facing the SC transportation industry, with a focus on the use of emerging information and data-driven technologies (referred to as the five driving emerging technological forces) for SC transportation applications. The paper also examines several use cases that can be exploited for SC transportation strategies.

In this paper, we present a comprehensive review and representative studies with the focus on the emerging technologies from five information and data-driven perspectives: (1) geoinformation approaches; (2) data analytics approaches; (3) machine learning approaches; (4) integrated deep learning approaches; (5) artificial intelligence approaches. The remainder of the paper is as follows. Section 2 presents the research method while Section 3 gives an overview of smart transportation and comparison. This is followed by Sections 4–6 which give discussions on the impacts of geo-information on SC transportation, data-driven transportation, and big data technology, as well as machine learning approaches for SC transportation. Section 7 discusses recent trends using integrated deep learning towards SC transportation and Section 8 gives some discussions on transportation empowered by other artificial intelligence (AI) techniques. Section 9 concludes the paper.

2. Research Method

One of the main objectives and contributions of this paper was to present a comprehensive study of the state-of-the-art for SC transportation techniques and approaches, with a focus on the emerging technologies, termed as the five driving technological forces. The literature review has been considered as a valid approach and a necessary step in structuring a research field, and thus constitutes an integral part of research [16]. Consequently, this study adopts a four-step research method used in [16,17] for collecting and analyzing the literature, namely, (1) defining the unit of analysis, (2) selecting the classification context, (3) collecting publications and delineating the field, (4) analyzing or evaluating the materials. Thus, within the parameters of this objective, this study presents literature covering over 867 research articles from journals and over 203 cited references at the end.

The literature has been structured under two main categories, namely: (1) an overview of smart city transportation and comparison, and (2) emerging technologies for SC transportation. Figure 2 shows a summary of the scope of reviews in this paper and Table 1 shows a summary of the classification descriptors and references showing areas, discussion, and studies. The relevant papers have been searched using Google Scholar, IEEE Explore, and Scopus databases from 2010 onwards, while others have been obtained via cross-referencing. However, a wide range of publications was found between 2017 and 2020, as shown in Figure 3; this was instrumental for the authors to refine their search in order to identify any missing publications, including traditional and new items relating to the keywords. In finding relevant publications, the authors also considered a single research paper as the unit of analysis, and employed a set of keywords to ensure the collection of a large number of studies. These studies have been analyzed under two major contexts, namely: (1) the problem context and (2) the solution/methodology context, to sufficiently cover both studies on smart city transportation strategies, as well as methods utilizing emerging technologies for smart city transportation applications, in order to address problems.

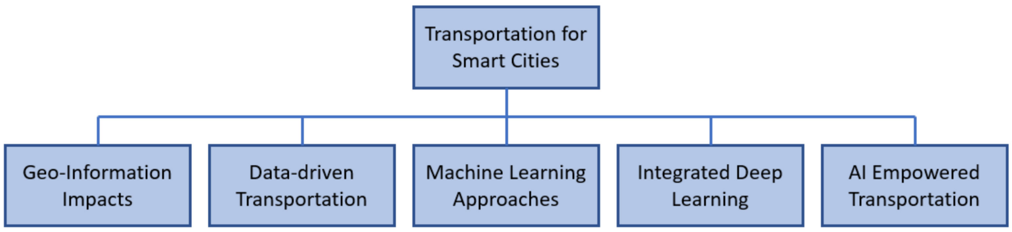

Figure 2. Scope of reviews.

Table 1. Classification descriptors and references.

Classification Descriptor	References
Overview of Smart City Transportation and Comparison	
Intelligent transportation	[18,19]
Transportation system architectures	[20–24]
Traffic monitoring and management	[25–28]
Social transportation and crowdsourcing	[29–31]
Platooning for sustainable transportation	[32,33]
UAV-enabled transport for smart city	[34]
Ridesharing in smart city	[35]
Multi-station vehicle sharing in smart city	[36]
Waste transportation in smart city	[37]
Emerging Technologies for Smart City Transportation	
Impacts of geo-information on smart city transportation	[38–66]
Data-driven transportation and Big data technology	[67–136]
Machine learning approaches for smart city transportation	[137–159]
Integrated deep learning towards smart city transportation	[160–182]
Transportation empowered by AI and other techniques	[183–205]

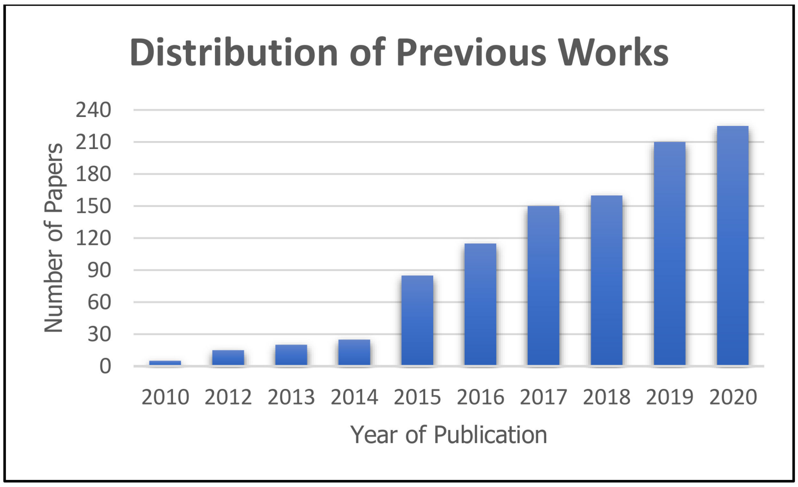

Figure 3. Statistics of distribution of the papers by the publication year.

3. Overview of SC Transportation and Comparison

This section presents an overview of SC transportation and the comparison of some of the literature. Several research efforts have been made to study previous works and determine how the SC transportation-enabling technologies can effectively be utilised. Smart cities (SCs) utilize a variety of sensors for data collection to perform analytics. The authors in [18] discussed several characteristics of SCs such as smart governance, smart communities, smart economy, smart environment, and smart mobility. The characteristic of smart mobility refers to the integration of ICT and sustainable transportation. The role of ITS in SCs is a key focus of a SC. A smart transport system allows for several useful applications, such as traffic monitoring, and informs traffic participants about potential hazards and situations. The following discussions give an overview of SC transportation, which are compared with the current review.

Intelligent Transportation: Intelligent transportation systems (ITS) utilize various technologies ranging from core applications such as traffic signals, control, and monitoring systems to useful applications such as parking guidance and decision-based information systems. An example of an intelligent transportation application is the intelligent vigilare system (IVS) for intelligent transportation services for SCs proposed by the authors in [19]. Figure 4 shows the architecture of the proposed IVS framework. The IVS utilizes several forms of ICT technology for data sensing, information processing, and cloud storage, which are integrated to perform the useful IVS application. The authors proposed the application of one of the five driving technological forces under our review—big data technology (incorporating IoT, ICT, and data-mining technologies, which they termed as the three pillars of any SC transportation project) for ITS. Their main purpose was to utilize the big data technologies to tackle the various ITS issues, including safety, security, and management. In contrast, we studied the combination of the five driving technological forces in the context of SC transportation strategies. Our main purpose was to acquaint researchers with the recent trends using integrated deep learning approaches and with useful insights into the data-driven approaches that can be exploited for SCs and transportation architectures.

Transportation System Architectures: This section presents some review studies on SC transportation architectures. The authors in [20] conducted a study and discussed various transportation architectures for SC applications. The architectures addressed different applications, such as traffic communications, shared vehicles, navigation, and energy. In [21], the authors proposed a cloud-based, smart car-parking transportation architecture for SCs. They first established this in a university campus using three tiers in the architecture (see Figure 5): (1) Cloud Tier—cloud storage and computation services. The data include parking lots, the car driven, and the user's location; (2) Web Servers Tier—connection of the mobile applications tier and the cloud tier. It supports the deployment of applications and provides the environment to modularize the applications into a bundle; and (3) Mobile Applications Tier—makes requests to the car parking web server asking for available car parking lots. The server will find an available car parking lot by following the user's profile and return application driving information to the user.

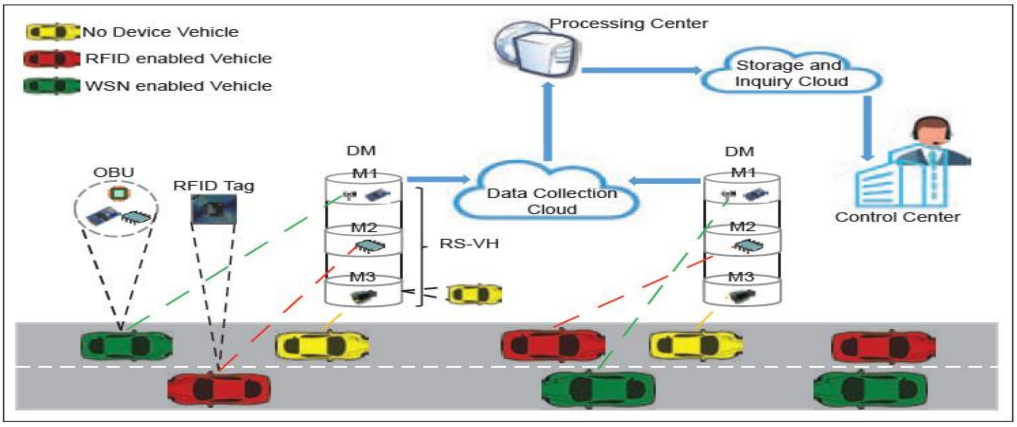

Figure 4. Architecture of IVS framework [18].

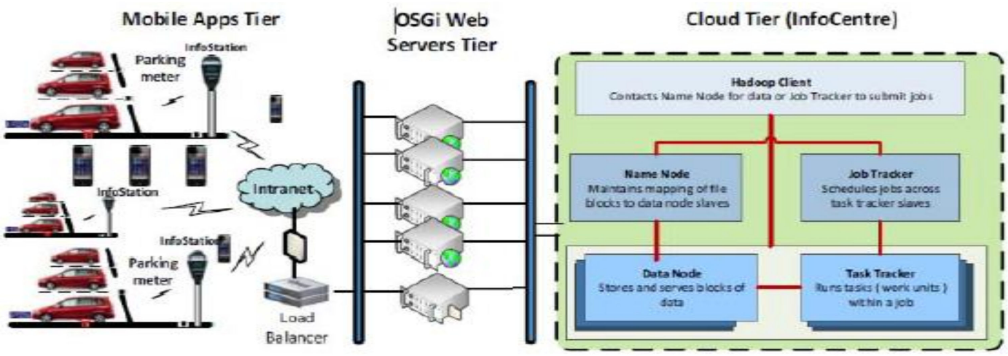

Figure 5. Smart parking transportation architecture [21].

The authors in [22] proposed a smart public transport system architecture by utilizing smart GPS-based buses, smart ticketing, and automatic fare collection. Their system architecture consisted of three modules: (1) Bus Transmitter Module; (2) Bus Terminus Control Module; and (3) Passenger Service Module. The Bus Transmitter Module consisted of a GPS module, a single board computer (Raspberry Pi), and a GSM Module. The Passenger Service Module consisted of a smart phone application to find the location of the bus and the time it will take to reach the destination. The authors in [23] discussed a SC transportation system and a framework for systems governance using the classification of the Singapore Land Transportation System (SLTS). Figure 6 shows the entities and relationship within the SLTS. The governance framework for the next-generation SC transportation system is shown depicting the governing body and the SLTS. The authors in [24] proposed an approach termed as ACP (artificial system, computational experiment, and parallel execution) and focused on a parallel transportation management and control system (PTMS). A new architecture for building new generation intelligent transportation systems (ITS), which is an expansion of PTMS, was proposed. Figure 7 shows the proposed ACP smart transportation architecture. Their approach utilized IoT and cloud-computing technologies for social transportation and agent-based systems. In their approach, the agent-based traffic control utilizes its autonomy and adaptability to handle the dynamic nature of the traffic environments. In contrast to conventional control approaches, which are static, the agents can be adapted in real time based on the surrounding traffic status.

the network. The application was implemented on the MapReduce framework.

The traffic agents are implemented as software as a service (SaaS) and are mobile within the network. The application was implemented on the MapReduce framework.

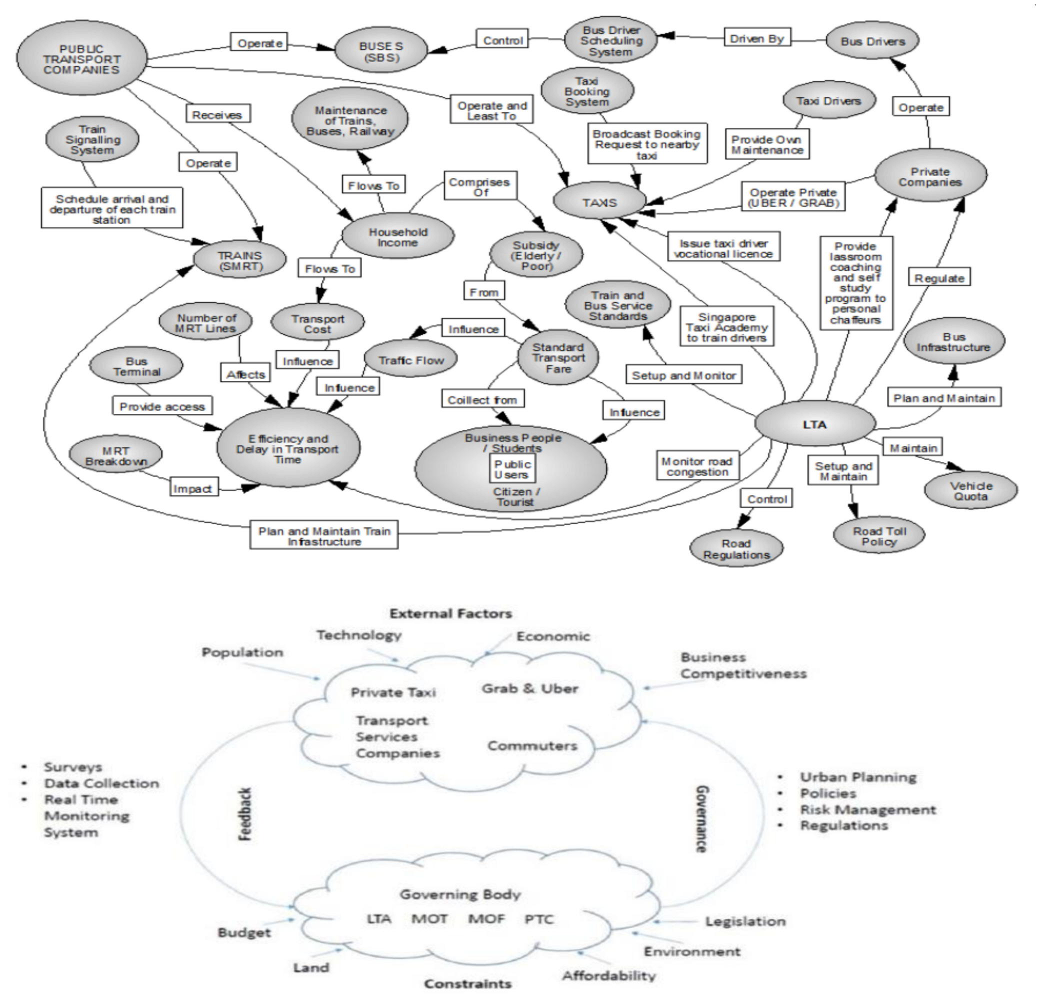

Figure 6. SLTS main entities and relationships [23].

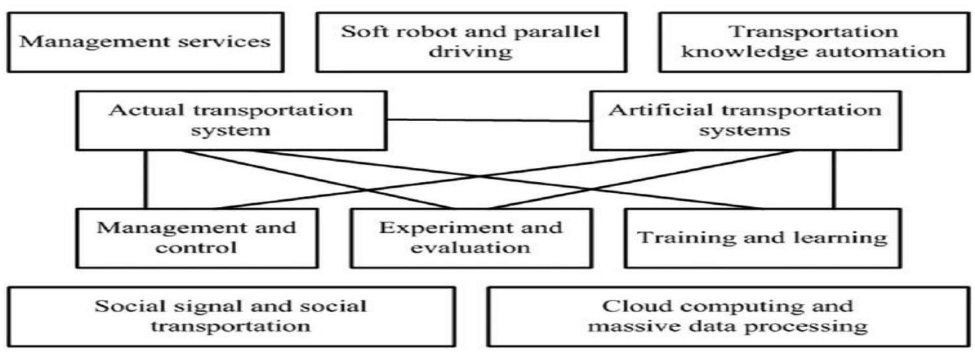

Figure 7. ACP smart transportation architecture [24].

The above discussions are limited to existing SC transportation architectures and compared to various parameters with the aim of building robust and generic software architectures for the SC transportation sector. The proposed architectural solutions are targeted to address a few issues, including car-parking, navigation, energy management, etc. When compared to our study, the aforementioned discussions do not present a comprehensive review study and are limited to only architectural solutions. Moreover, recent studies reveal that the challenges of the SC transportation industry surpass the mere provision and application of only architectural solutions, but rather require some integrated solution–strategies, which are examined in the present study. This is because such architectural solutions are not encompassing and have a limited capacity to address a variety of issues affecting the SC transportation sector.

Traffic Monitoring and Management—This section reviews a few research works related to traffic management and monitoring for SCs. The objective of traffic management is to ensure the effectiveness of intersections, roads, and motorways. These systems provide useful information to road users (e.g., real-time information and traffic density forecasts) and on the implementation of intelligent systems to reduce negative impacts (e.g., road accidents, traffic congestion). Useful information on traffic patterns can be gathered by connecting smart lighting and signals with traffic control systems. The authors in [25] proposed an intelligent traffic system to reduce the waiting time for vehicles. Their adaptive system uses detectors to gather information about the state of the road, which is used to calculate the optimal traffic signal time. The authors in [26] proposed an approach for spatio-temporal congestion-aware path planning for ITS. They embedded SDN technology into the ITS and proposed a grid-based model to quantify the traffic-congestion probability for forecasting. Figure 8 shows the SDN-enabled SC and the grid-square-embedded model. The traffic-congestion probability of an area is proportional to the average traffic flow at each square area.

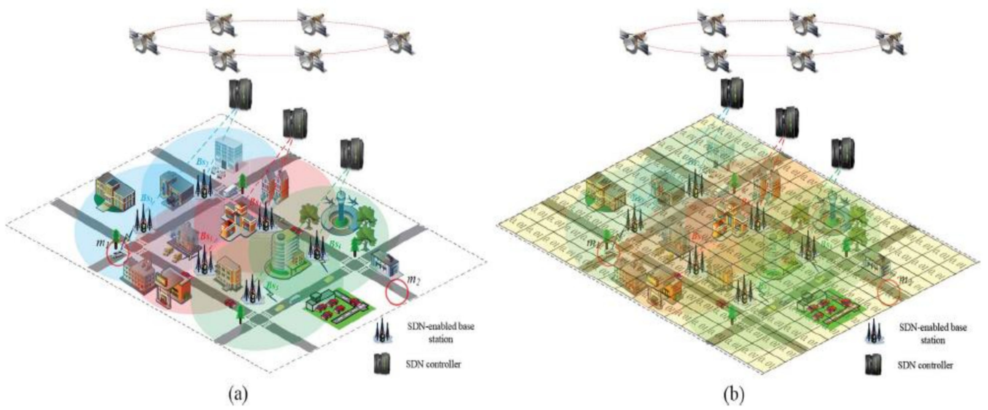

Figure 8. SDN-enabled smart city and the grid-square embedded model [26]: (**a**) SDN-enabled SC with distributed path planning management; (**b**) Grid-based network model.

Traffic flow prediction is an important problem to address for SC environments. The predictive tasks can be classified into three categories based on the traffic duration to be forecasted: (1) long-term traffic prediction; (2) medium-term traffic prediction; and (3) short-term traffic prediction. The authors in [27] have presented a review in this SC application area. The authors in [28] classify traffic flow prediction approaches into five categories: (1) statistical analysis models; (2) artificial intelligence (AI) models; (3) nonlinear theory models; (4) traffic simulation models; (5) combined prediction models. The aforementioned studies are restricted to traffic flow predictions and monitoring, and thus have limitations with regards to devising comprehensive solution–strategies, which require a collaboration of the five technological forces, as presented in our study for SC transportation applications. The summation of their investigations is just a component of the several solution mechanisms reviewed in the present study. In contrast to our study, their studies lack a comprehensive review of an integrated approach to address the various issues affecting the SC transportation industry.

Social Transportation and Crowdsourcing—Social networking systems can be utilised to develop smart transportation systems. For example, maintenance and improvement of public transportation services are important in SCs. However, the implementation of new and improved features can be costly. A social-based approach can be utilised to collect real-time tracking data using participatory sensing, also termed as mobile crowdsensing or crowdsourcing. In this approach, the mobile devices from passengers are used to collect data. The authors in [29] proposed a crowdsensing-based public transport information service for SC applications (shown in Figure 9).

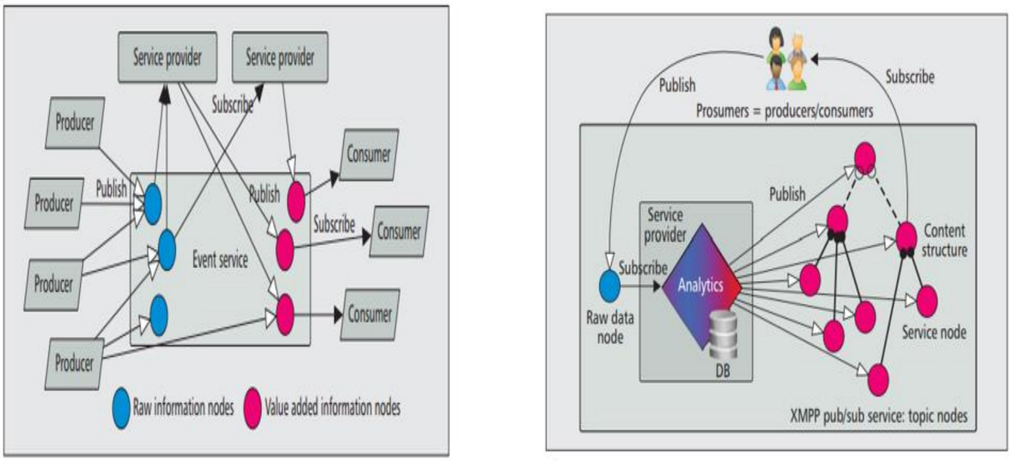

Figure 9. Mobile crowdsensing smart transportation architecture [29].

Their approach utilized an Android user interface termed as TrafficInfo (shown in Figure 10), which is implemented on a XMPP communication framework to facilitate the development of the crowd-assisted smart city application. The authors used XMPP and its generic publish/subscribe communication model in the framework to implement interactions. It could also be configured to use the OpenStreetMap (OSM) [30], which is a crowdsourcing-based mapping service. The TrafficInfo has three main features: (1) visualization; (2) information sharing, and (3) sensing. The figure shows the TrafficInfo screenshots for vehicle visualization, the user feedback form and the sensor data flow in TrafficInfo. Again, these approaches do not address the critical issues of the SC transportation sector, as being examined in our study, which require an integrated solution using the five emerging technologies currently being reviewed by the present study.

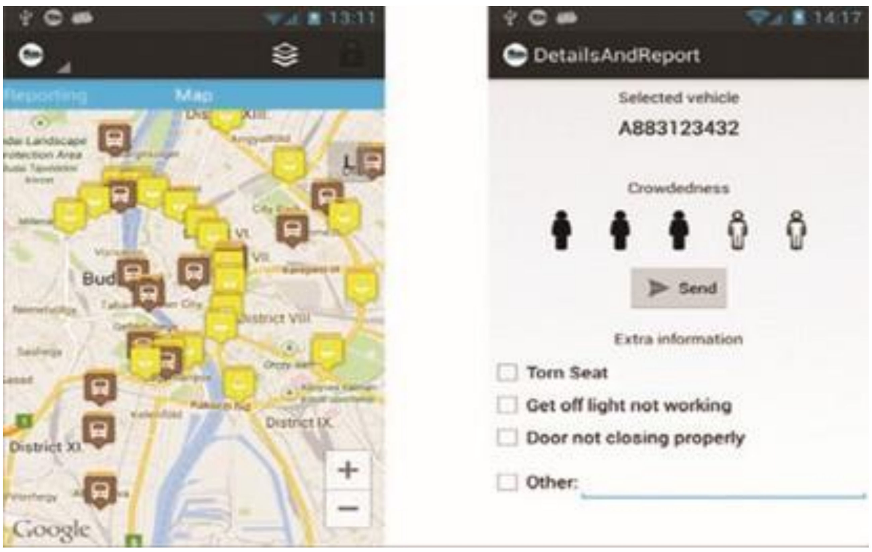

Figure 10. TrafficInfo smart transportation application [29].

Package delivery services are in high demand due to the rapid growth of online retailers. The authors in [31] proposed a city-wide package distribution and a framework called the crowdsourced public transportation system (CPTSs). This approach aims to utilise the of idle capacity of CPTS vehicles. There are four states (waiting, riding, re-waiting, and unloaded) in package delivery. The delivery scheme determines the state of any package at any time slot, and calculates the optimal time for delivering all packages. The authors formulated the problem as an NP-hard problem and proposed an efficient heuristic solution using ILP techniques. Figure 11 shows a comparison between the traditional logistic distribution model, where each logistic company distributes packages to their customers independently, and the proposed approach. Their experimental work was validated using simulations with data from a real bus transportation network.

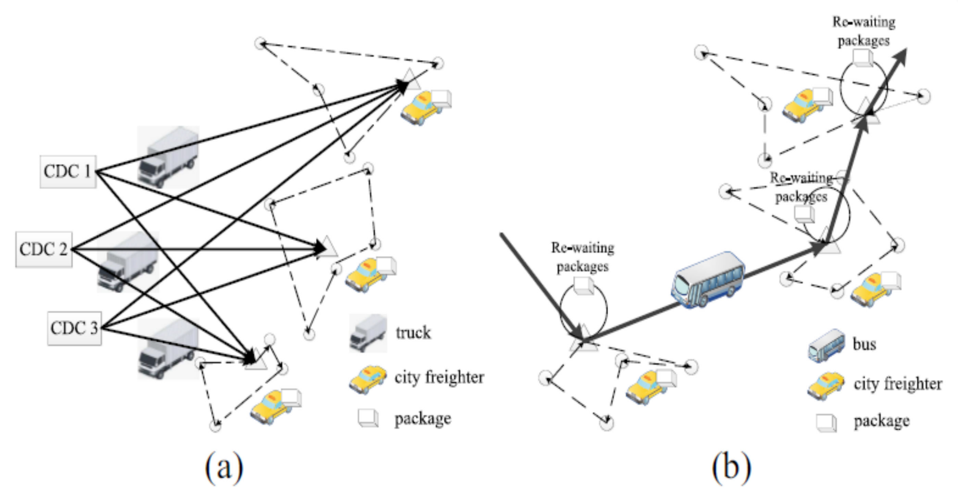

Figure 11. Comparison between traditional logistic distribution (**a**) and CPTS model (**b**) [31].

Platooning for Sustainable Transportation in Smart Cities—In an SC, it is important to ensure the sustainable development of urban transportation, which considers fuel consumption as well as traffic efficiency. Platooning is a cooperative driving application where autonomous vehicles move on the same lane in a train-like manner [32]. The vehicles in the lane are maintained at a constant inter-vehicle distance to reduce fuel consumption and to achieve road safety. The authors in [33] proposed a systematic framework for vehicular platoon formation. In their approach, an optimal speed model is first computed to optimise the total fuel consumption. Next, an insertion point based on a Q-learning model is derived and used for the vehicles in the platoon. Their approach also designed a collision detection model for new vehicles joining the platoon.

UAV-Enabled Transport for Smart Cities—UAVs have also been proposed to play a role in smart transportation for SCs. The authors in [34] discussed the potential of UAVs in SC transportation. Some examples of the applications that can be enabled by UAVs include aerial accident report agents, aerial speed cameras, aerial policing, and aerial traffic signals. Figure 12 shows a UAV smart transportation architecture for an aerial accident report agent, where one UAV could fly to the accident location and issue a report or alarm (e.g., video), and then land and transmit its report or alarm for other UAVs.

Ridesharing in Smart Cites—Ridesharing in smart cities provides benefits such as a reduction in traffic congestion, reductions in carbon footprint and travel cost, enabling a partial solution to parking problems, etc. The authors in [35] discussed the possibility of quantifying an individual contribution towards sustainability and a reciprocal incentive approach to encourage voluntary behavioural development towards sustainable mobility solutions. They also proposed a framework called WeDoShare, which is a ridesharing framework in transportation for sustainable mobility in SCs. Figure 13 shows the conceptual view of WeDoShare in an IoT-based transportation System. This work addressed the issues of ridesharing and promoting the long-term engagement of Single Occupancy Vehicles (SOV) owners in ridesharing.

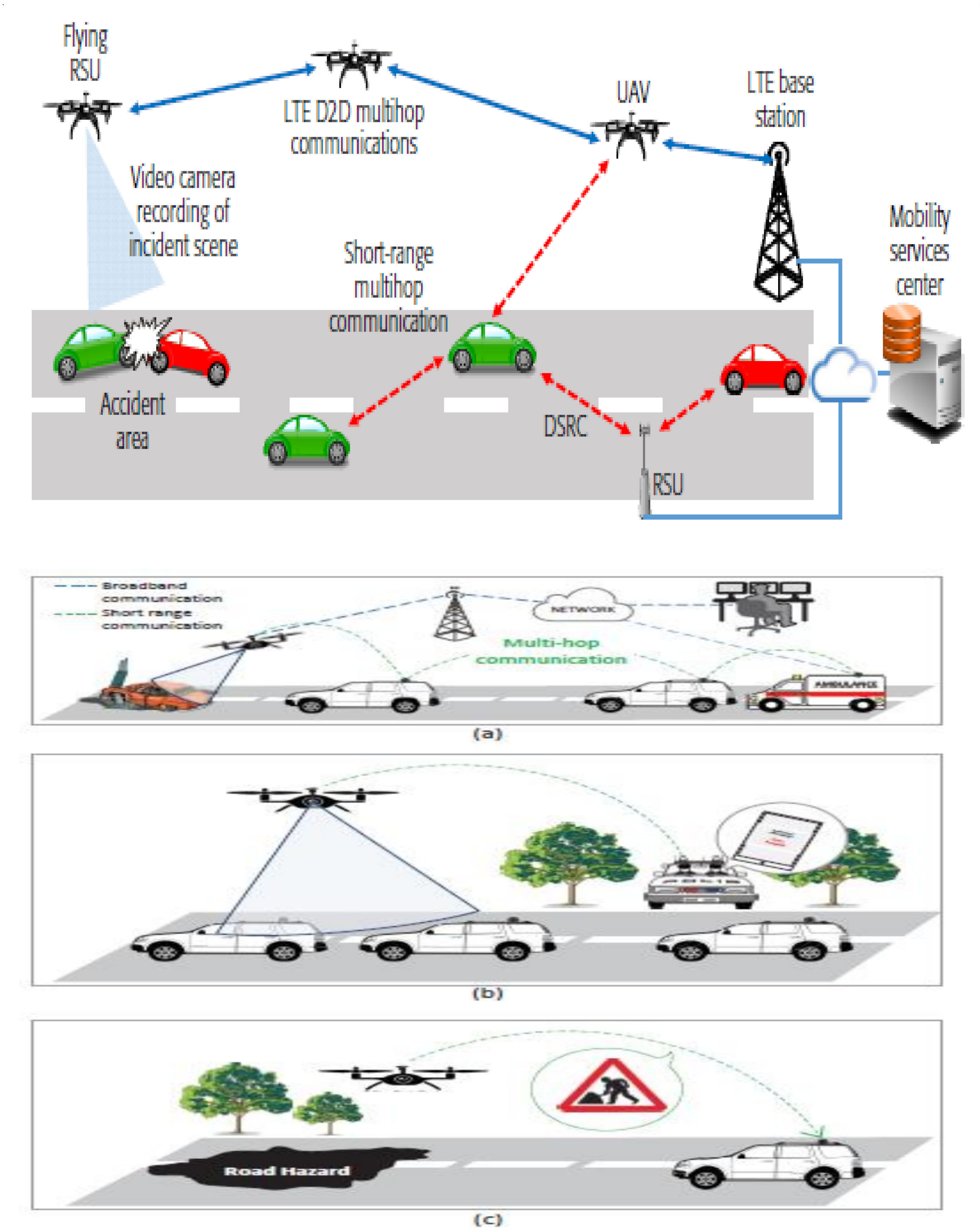

Figure 12. UAV smart transportation architecture [34]: (**a**) a UAV is exploited to provide the rescue team an advance report prior to reaching the incident scene; (**b**) Police exploits UAV to catch traffic violations; (**c**) a UAV is used as a flying RSU that broadcasts a warning about road hazards detected in areas devoid of an RSU.

Multi-Station Vehicle Sharing in Smart Cities—The authors in [36] proposed a transportation system architecture for multi-station vehicle sharing. There are three components in this architecture: (1) the User Trip Registration component; (2) the System Management component; (3) the Vehicle component. The User Trip Registration Component registers the requests for vehicles, which requires signing up. The System Management Component contains the information of users, vehicles, and requests. The data from the vehicles and registration kiosks can be analyzed to give information on user trip behavior and vehicle operation. Radio transponders can be used to communicate with the system management.

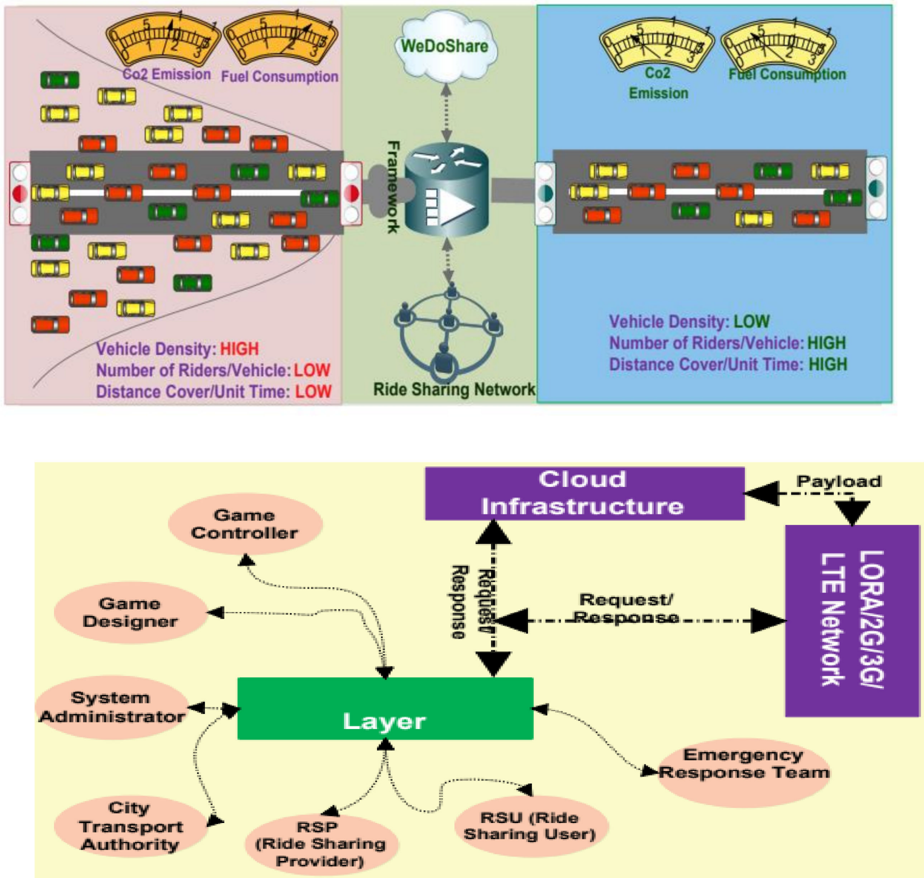

Figure 13. Ride sharing transportation architecture [35].

Waste Transportation in Smart Cities—An important function in SCs is solid waste collection and management. The authors in [37] discussed a study on methods for efficient waste transportation and recycling. Their work considered IoT-based solutions for waste transportation management, which included two use cases: (1) survey and data collection; (2) IoT-based smart waste transportation system. The authors performed simulations and implemented a prototype system using a case study of a metro region. The simulation scenarios were for: (1) waste collection; (2) waste recycling. Figure 14 shows the waste collection simulation and the waste recycling simulation. Their work demonstrated several benefits for SCs, including reducing traffic congestion, and thus saving both fuel consumption and time.

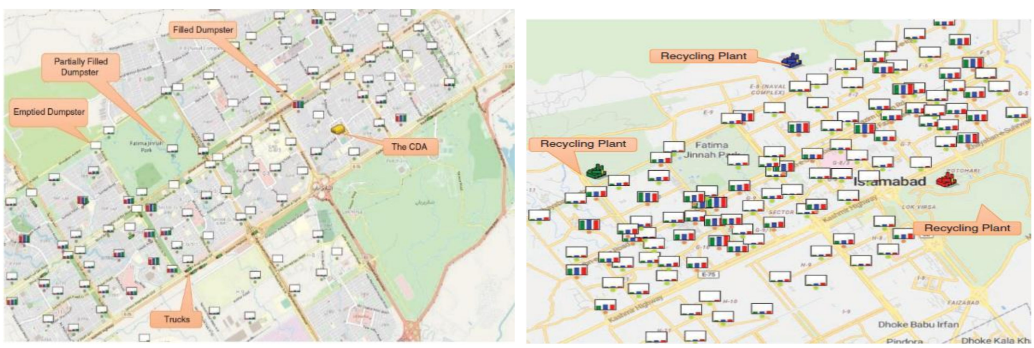

Figure 14. Waste transportation architecture [37].

All the aforementioned works are advancing the topic but have overlooked or oversighted the need to comprehensively examine the technological impacts of the five driving forces and emerging technologies—(1) geoinformation, (2) data-driven & big data technology, (3) machine learning, (4) integrated deep learning, and (5) artificial intelligence (AI) in the context of SC transportation strategies. Such a comprehensive study is crucial and critical with regards to acquainting researchers with the recent trends and state-of-the-art technologies that can be exploited for SC transportation strategies. When compared to other approaches, this study aimed to comprehensively examine the combined impacts of the five driving technological forces in the context of SC transportation strategies. We also note that, to the best of our knowledge, this is the first comprehensive review to consider and examine these five technological approaches for SC transportation applications.

4. Impacts of Geo-Information on SC Transportation

This section presents the impact of geo-information on SC transportation and some representative works in this area. Some studies on the utilization of geoinformation and geospatial technologies in SCs can be found in [38–40]. The authors in [41] discussed the transformation of traditional urban spaces to digital spaces and the utilization of geoinformation for deployment in SCs. The authors in [42] explored SC frameworks and the environmental impacts of ICT solutions. Geoinformation and communication technology (GeoICT) offers solutions for sustainability in SCs. The authors in [43] presented a study which highlighted the role of GeoICT in the development of SCs. The authors in [39] presented an approach for geomatics contribution called GeoSmartCity. Their approach utilized data such as location, routing, correlation, and spatial interactions. Visualization tools provide real-time information on colored maps showing traffic flow.

The authors in [44] discussed the application of GeoDesign in SCs and how three-dimensional planning can be utilized for sustainable design. The authors in [45] proposed an approach to three-dimensional model fusion in geographical information systems (GIS). In this work, the authors discussed the usefulness of a three-dimensional model which can be used as an input for other urban models such as telecommunication networks, disaster management, and renewable energy planning. The authors in [46] proposed a multi-scale 3D-GIS approach for the assessment and dissemination of solar income of digital city models. The authors in [47] proposed the usage of location-based services (LBS) by the utilization of mobile devices, networks, and service providers. The location of users and their contexts are important to LBS [48]. Some applications of LBS for SCs include security, disaster management, and mobile workforce management.

The authors in [49–51] proposed utilizing large volume of spatiotemporal and big data for SCs. Location analytics or geospatial big data analytics [52] involves the use of geo-computing and spatial analysis to mine for new knowledge which could be embedded in the spatio-temporal data. Examples of applications can be found in the analysis of geo-tagged location data (e.g., Twitter), such as activity patterns, population estimation, and disaster management [53–56]. Advancements in remote sensing (e.g., satellite data) could also assist by providing information about the environment [57]. These GeoICTs are derived from satellite images and form the concept of the digital earth and a georeferenced digital representation of the planet. The use of geospatial technologies can be utilized towards analysis and monitoring of urban areas based on environmental indicators [58].

The authors in [59] proposed an approach to analyze vulnerabilities in urban transport systems through the metrics of complex networks. Complex network metrics is an approach to model complex interactions between objects in high dimensionality (e.g., transportation routes and stops can be modelled as a graph). The authors in [60] discussed some key technologies in smart transportation for SCs. Figure 15 shows the structural diagram of the smart transportation system. The system utilizes different ICT technologies to combine the interactions and communication among the various entities (e.g., people, vehicles, roads) into a digital map. Some applications include the selection of the optimal driving route based on real-time traffic information and location of vehicles, using the GPS positioning technology.

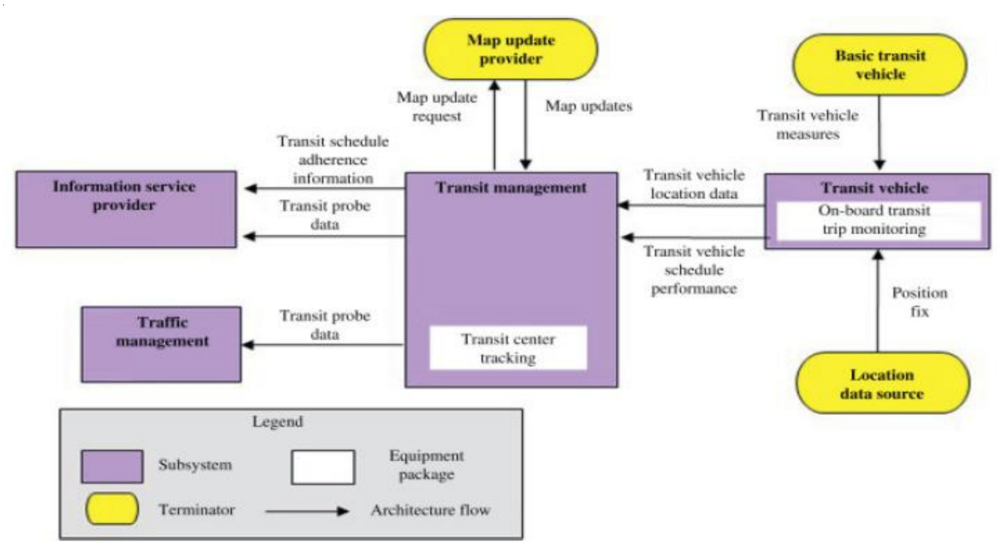

Figure 15. Structural diagram of smart transport architecture [58].

The authors in [61] proposed an approach termed as ELTRO for a geo-information model of a magnetic levitation transport system. A recent advancement in GIS presents the concept of simultaneous localization and planning (SLAM). The authors in [62] proposed a framework to integrate the GIS and SLAM technologies for smart transportation systems. Their work proposed the utilization of SLAM for power distribution networks to support urban transport systems and the charging of electric vehicles (EVs). Their implementation utilized a cloud-based centralized SLAM on the GIS framework to ensure stable power delivery, based on load forecasting through vehicle localization. The authors in [63] proposed geostatistical methods for estimating individual transport speeds. Their approach utilized geographically weighted regression techniques and methods of Kriging family and considered real urban conditions. Their experimental work used real data from Novosibirsk. The authors in [64] proposed an urban sprawl classification analysis using image processing in GIS. Their work used Lansat satellite images for feature extraction.

The authors in [65] proposed an approach termed as the latent factor modelling of traffic trajectory data. They proposed a generative model termed as TraLFM to mine the human mobility patterns underlying traffic trajectories. Their approach used three observations: (1) mobility patterns are reflected by the sequences of locations in the trajectories; (2) mobility patterns vary with people; (3) mobility patterns are cyclical and vary over time. The authors performed extensive experiments on vehicle passage records (VPR) and taxi data. The VPR data were collected from the traffic surveillance system in the city of Jinan. The taxi data were collections from the complete trips of 442 taxis in the city of Porto. Their experimental results showed that the proposed TraLFM approach could outperform the predictions from other state-of-the-art methods.

The authors in [66] proposed an approach to estimate the influence distance of bus stops using the bus GPS data and bus stop properties. The influence distance of a bus stop is the distance between the start and end points of the zone where the road traffic nearby is adversely affected by the buses. Their work aimed to analyze the traffic behavior of the buses around the bus stops and to show the effects of the surrounding network. Their work also aimed to determine whether different traffic behavioral dynamics are influenced by the contextual properties (e.g., number of lanes, traffic light locations). Figure 16 shows the data preprocessing with map matching. Their experimental work was validated using data from 12 distinct bus routes in the city of Istanbul, which contained GPS data from over 5000 bus vehicles and 450 bus stops.

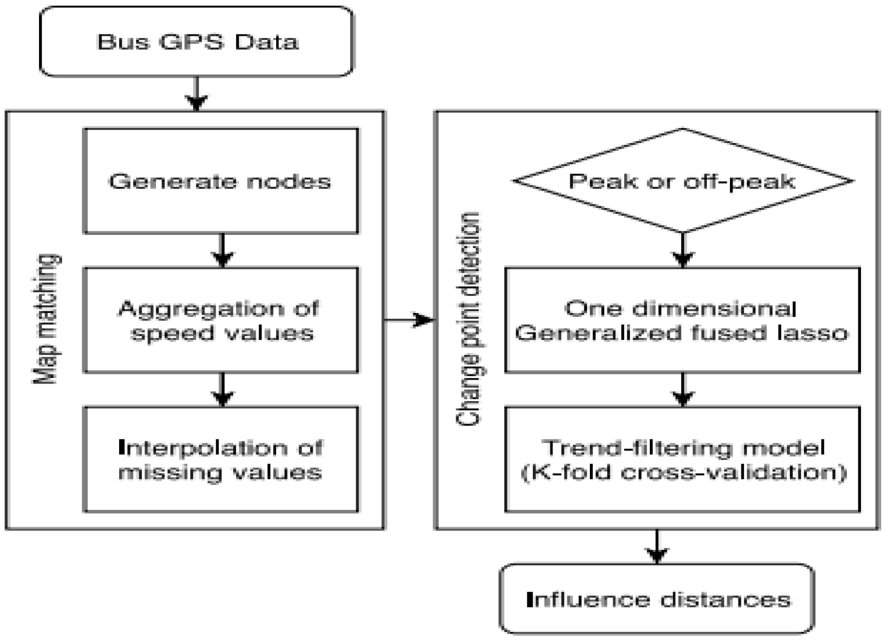

Figure 16. Data preprocessing with map matching [66].

5. Data-Driven Transportation and Big Data Technology

Data is revolutionizing our world and transforming the way we live in and interact with our society. It is transforming the way our society and its infrastructures are operated, and how industries produce goods and offer services. Data are considered as a digital commodity, which, if properly utilized, can pilot significant social change, as well as economic development [67]. With the development of SCs, massive number of sensor devices, wireless communication devices, hardware devices, as well as software applications have been integrated with infrastructures and services to generate data about our everyday life, which are utilized to enhance infrastructural development and service delivery.

The transportation system as an integral part of the society and a key element in the SC ecosystem generates large volumes of data daily. As a result of the volume, variety, and variability of transportation data, it is highly necessary to realize data-intensive jobs such as data integration, data visualization, data querying, and data analysis for extensive real-time applications, in order to efficiently utilize the potential of data. Existing conventional data processing tools are deficient in the effective realization of the aforementioned tasks for extensive real-time applications [68]. Big data in this context, has been proposed as the key technology that can efficiently address data-related challenges in the transportation industry, due to its ability to obtain, store, manage, and analyze large volumes of data to extract useful information needed for implementing ITS [69].

5.1. Data Collection and Sources

Advances in technology such as IoT, social media, etc. have widened sources through which huge amounts of data can be generated and collected from devices, vehicles, sensors, and people, for utilization in ITS. Data collection in ITS therefore refers to the job of collating data generated from these sources using suitable various channels. The authors in [70,71] provide insights on the transportation data sources and types. One of the transportation data collection sources is the smart card. Many urban transportation systems (BRT, rail, etc.) make use of automatic fare collection (AFC) systems, which require that passengers would have to use their smart cards to pay for fares. When passengers use (touch) their smart cards, the electronic reader captures details about their boarding time, source–destination information etc. Figure 17 shows some sources of data collection for ITS. Table 2 presents a summary of transportation data collection sources and their applications.

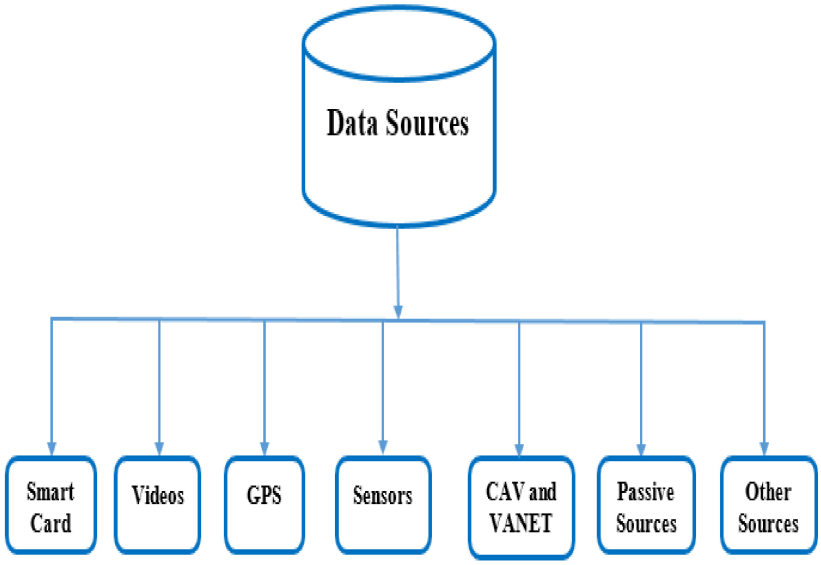

Figure 17. Data collection sources for SC transportation strategies.

Table 2. Summary of transportation data collection sources and applications.

Data Sources	Data Collection Tools	Data Type	Applications	Ref.
Smart card	Smart card	OD flows, travel time	Passenger travel behaviour pattern, public transportation services planning and management	[70]
Video	Video camera	Vehicle position, vehicle speed, vehicle density, vehicle classification, plate number	Traffic flow detection, and monitoring, vehicle identification, incident detection, vehicle emission modelling	[70,72]
GPS	GPS	Vehicle position, vehicle density, vehicle classification, road quality	Navigation services, traffic monitoring, travel mode detection, travel delay measurement, routing optimisation,	[73]
Road site sensors	Pneumatic road tubes, inductive magnetic loops, microwave radars, piezoelectric loop arrays, ultrasonic sensors, acoustic sensors, IR sensors, light detection and ranging (LIDAR) sensors, toll plazas, vehicle detectors,	Vehicle position, vehicle flow, vehicle density, vehicle speed, vehicle classification, trip time	Vehicle counting, and identification, traffic density and speed estimation, congestion prevention and route planning, short-time forecasting, parking demand modelling	[74]
Floating car sensors	Automatic vehicle identification (AVI), transponder, license plate recognition (LPR)	OD flow, travel time	Travel route selection and estimation, driver monitoring, driver behaviour modelling	[75]
Wide area sensor	GPS, cell phone tracking, video processing sound recording, photogrammetric processing, space-based radar, airborne sensors,	OD flow, travel time	Wide area traffic monitoring	[71,75]

Table 2. *Cont.*

Data Sources	Data Collection Tools	Data Type	Applications	Ref.
Connected and autonomous vehicles (CAVs) and VANET	Various sensors	Coordinates, speed, acceleration, safety data	Online vehicle diagnosis, smart charging planning, travel delay reduction, safety performance enhancement, congestion and accident detection, traffic flow prediction	[76–78]
Passive collection	Social media (tweeter, Facebook, Weibo,), mobile phone, collaborative applications	Travel time, OD flow	Real-time congestion avoidance routing, OD estimation	[79–81]
Other sources	Smart grid, smart meters, cellular service, dedicated tests	Electric and energy consumption, location, channel data	Performance and efficiency improvement, dashboard analysis	[82,83]

5.2. Data Analytics for Transportation

Data analytics is the process of inspection, selection, transformation, governance, and representation of data, with the goal of discovering useful information and assisting or boosting decision-making systems. Transportation is an integral part of our contemporary society. It provides the means of movement of people and goods and drives economic development. As the urban population surges, the transportation industry faces the challenge of developing effective and efficient strategies to utilize available infrastructure and reduce traffic [84]. With the advent of SC and ITS, sensors, GPS devices, smart phones, automatic fare collection (AFC) systems, etc., have been deployed to generate data daily from passengers, infrastructures, and transport services. The large volumes of data generated are processed with data processing models to gain valuable information that can be utilized by governments and private companies to conduct in-depth analysis and monitoring to make informed decisions to enhance the quality of their services.

However, most conventional tools for data analysis, such as relational databases, are deficient in storing and managing such highly voluminous data [85]. Recently, big data analytics has been applied in many sectors including the transportation industry because it provides big data platforms at the highest level for the application of learning, pattern inference, and optimization techniques. Big data provides some features that give it a competitive edge over traditional models. Several studies propose that applying a simple model to highly voluminous data produces a more reliable accurate result than applying sophisticated learning algorithms to less voluminous data [68]. Also, big data analytics defines the sophistication of the models themselves and how they support scholars obtaining fresh insight. Big data promises speed and efficiency in the transportation services, although the enormousness of real-time data in transportation, and in the manner that big data analytics predicts the condition of components in ITS, constitute some major challenges that need to be addressed. In this context, several studies have been conducted to incorporate big data analytics into the transportation sector.

The authors in [87] designed an analytic technique to manage the massive data regarding traffic volume count, and to carry out congestion analysis. The study was done using huge traffic volume count data collected by radio frequency identification (RFID) devices in Nanjing, China. Their aim was to apply software, analytical, and virtualization approaches to data to determine the peak hours, off-peak hours, and places of high and low traffic volume. The authors acquired three sets of data for vehicle movement, RFID reader, and vehicle type. Attributes captured in the vehicle movement data included license number, the time and date detected by the RFID reader, the vehicle type identification code, the road lane's number, where the vehicle was detected, and the RFID reader identification codes that detected the vehicle.

The authors in [88] proposed a real-time vehicle traffic analysis framework based on big data analytics and IoT devices to provide predictions for future policy-making in smart transportation. The study was carried out in four routes in the Ernakulam district of the state of Kerala during a two-peak period of 7.00 am to 8:30 am and 4:00 pm to 6:30 pm, using four buses. Their aim was to provide the decision-making authorities a guide for developing transport policies, such as deciding if a road bumper or speed breaker is needed by correlating the data with daily traffic analysis, and in allocating funds to infrastructure projects in the future. The proposed system was composed of four modules—a GPS module, an IoT module for GPS data collection, a server module (for receiving data), and a mobile app module. The proposed system also employed Hadoop and Apache Spark for the analysis of stored historic data and real-time data, respectively. The GPS module was placed on vehicles while the receiver module was placed in the control section, and the data generated were collected and analyzed at various intervals of time to identify regular congested routes. The results were transmitted by the transmitter module to the control section, which, alongside with the software section, assesses the congestion and presents a graphical analysis of traffic flow as output to end users via a mobile application.

The authors in [89] proposed a method to estimate passengers alighting behaviors at bus stations using automated fare collection (AFC) transaction data generated from bus rapid transit (BRT) users' smart cards between bus terminals and corridors. The method used the origin–destination (OD) matrices as its parameters for estimation. The study used transaction data obtained from AFC payment systems, with the entry–exit mechanism of a BRT system in Jakarta. The data was preprocessed to remove duplicate data and single transactions, filter completed transactions (tap-in and tap-out), and classify transactions tapped at adjacent times. Information extracted and used for the algorithm implementation included the transaction timestamp, smart card serial no, flag status (in or out), bus stop, sub-corridors, corridors reference, and stops. The study estimated BRT user's alighting station in Jakarta at several levels of available corridors by developing OD matrices as a result. The result was validated with the AFC exit data as a valid alighting station. A comparison of the result of this method at different levels of validity with OD matrices estimation showed that this method gives results with up to 94% accuracy.

The authors in [90] employed taxi booking and GPS trajectories datasets combined with mobile-sensing datasets to build a system for city transportation service analytics called TRANSense. To give a practical realization of the proposed framework, the authors demonstrated two analytic applications: (1) the taxi service analyzer (TSA) which detects commuters queuing for taxis and utilizes taxi trip information to identify possible high demand taxi locations and selectively activate mobile sensing-based analytics for nearby commuters to evaluate their waiting time; (2) subway boarding analyzer (SBA), which identifies instances of travelers failure to board arriving trains. It achieved this by approximating the arrival times of trains from the temporal patterns of travelers' departure at station booths, and subsequently used mobile sensing-based analysis of travelers' behavioral movement on platforms. Experiments performed with real-world datasets collected from over 20,000 taxis and 1.7 million passengers in Singapore revealed that TSA detected passengers queuing with an accuracy of over 90%, with an insignificant energy overhead, and estimated wait time with less than a 15% error margin. On the other hand, SBA detected failed boarding situations with over 90% precision.

The authors in [91] applied data analytics to determine city hotspots depending on time and location, using the 2016 city of Chicago big taxi dataset. The authors provided descriptive, predictive, and prescriptive analysis to help determine taxi companies that needed improvements in customer service, maximize drivers' earnings, obtain information on trips, examine the average range of drivers' drop-offs and next pick-ups, and work out better commute patterns. In the descriptive analysis, the authors visualized the data by employing the Python library MatPlotLib to enable the plotting of the data to discover patterns. The graph revealed the volume of trips undertaken by the ten highest earning companies.

The authors in [92] employed real-time big data acquired from the Florida Department of Transport (FDOT), online data streamed from vehicles on the road, and data collected from vehicle detectors on the roadside to develop a big data system to increase roadway safety, prevent or minimize accidents, and decrease congestion. Their method was based on splitting the roadway into sections utilizing the available infrastructure and minor accident features. They applied a linear regression (LR) model to the data to accurately provide the estimated time of arrival (ETA), while naïve Bayes (NB) and distributed random forest (DRF) were used to predict accidents and congestion before they occurred. In an accident or congestion situation, the ETA is updated by predicting the correct time needed to clear it. To make the proposed system fast, accurate, and reliable, the authors integrated the lambda architecture into their framework, based on its ability to provide scalability, speed and fault tolerance. They ensured that a relevant set of features was selected to enhance the efficiency, accuracy, and speed of the proposed model.

Using city-scale transport data (bus, subway, etc.,), the authors in [93] developed an analytics framework for identifying tourists and understanding their preferences, called TourSense. They first presented a graph-based iterative propagation learning algorithm to identify tourists from public passengers and designed an analytical model for tourist preferences to learn and predict the tourists' next trips. Experiments performed on real-work datasets collected from over 5.1 million passengers and their 462 million trips showed that the proposed framework was very effective. In [94], authors designed a proof-of-concept system that analyses bus schedules and real-time bus locations for the city of Brasov, Romania. Their aim was to develop a system to revamp public transportation in the city and to enhance the city's attractiveness. The system was built atop of the CityPulse framework, which supports many smart city solution creations using data analytics, real-time IoT data, distributed systems, social media data streams, and so on.

The authors in [95] proposed a MATLAB-based data analytic method to model the demand and route planning for a bus transport system, utilizing data gathered from electronic ticket machines (ETM). The authors used ETM data generated from the daily ticket sales transactions made by the Road Transport Corporation in the state of Kerala (KSRTC), at six bus stations in the city of Trivandrum between the period of 2010 and 2013. The extremely huge dataset generated by the ETM is estimated to be around one million points, on average, of commuters' monthly transactions. The authors focused on auditing and compiling the dataset to ascertain commuters' demand, operators' performance, and operators' service effectiveness. With the MATLAB tool, the dataset could be queried to ascertain the origin–destination (OD) matrix of bus passengers, which assists in modelling demands, in making decisions, and in formulating policies for future predesign of the transport system. Likewise, it is possible with the use of the analytic tool to ascertain the link-volume of the transport network, and also to obtain information about boarding and alighting of commuters at bus stops from the ETM dataset.

The authors in [85] designed a data analytic approach integrated with Hadoop to enhance the operational efficiency of transportation and logistics companies. The authors collected data about speed, fuel, acceleration, driver's ID, date, GPS location coordinates and time from vehicle sensors and GPS devices, which were then transferred to a Hadoop clustered server. These raw data were generated and sent in packets to a HDFS system by hundreds of motor vehicles every two seconds. These terabytes of raw data are analyzed weekly or monthly using analytics method to enable transportation firms to monitor the driving behavior of drivers, determine fuel usage, and assess risks undertaken by drivers. This will help to cut down cost and enhance productivity.

The authors in [97] proposed an effective route planning technique called SubBus for shared subway commute buses using crowd-sourced mobile data to predict commuter flow at bus stations and for the effective planning of routes. First, the authors analyzed the travel behaviors of residents to ascertain five predictive attributes which included flow, location, time, week, and bus, and used them to predict the travel needs for shared buses using machine learning model. Using the operating features of the shared buses, the authors designed a dynamic programming algorithm to create dynamic a best route with fixed destinations for several running buses. The authors therefore incorporated the dynamic programming algorithm together with the five predictive features into the shared bus dynamic route planning approach (SubBus) to plan workable routes using the actively changing travel needs. The evaluation of SubBus showed it outperformed other methods.

The authors in [98] proposed a hybrid data-driven transportation simulation model that can assess and visualize parameters that measure network performance to enhance the operational response to real-world smart city scenarios. The traffic model incorporates the merging of preset data intersections and real-time data-driven intersections and depicts a traffic passageway partly provisioned with smart devices capturing high velocity, high volume datasets with short lifespans. The model has emulated seventeen successive intersections on a passageway where the vehicle volume and signal control at two intersections were driven by real-time emulation data from in-field sensors, while the other intersections were controlled by preset data in the calibrated model. The study used various Python and PERL scripts to fetch data continuously from the in-field sensors to populate the database. Because of the high volume, velocity, and broad variation of the stream of data, the architecture employed data analytics to extract useful features. The result of the simulation with the proposed hybrid model showed strong effectiveness in working with high volumes of data when compared to models that employed preset values. The proposed model also responded sufficiently to changing values of real-time input data.

The authors in [99] designed a graph-based algorithm to reduce the pressure on the University of Nebraska Omaha (UNO) shuttle transport system. The system utilized data about time, the number of students expected to use the shuttle, and the cost of fuel to optimize the route in order to reduce user pressure and minimize carbon impact of the transportation network. The implementation of the model when compared with an existing routing model revealed that an average of 1.2 min was saved per shuttle run. For 257 shuttle runs daily, the proposed model made a saving of 308.4 min daily and 1542 min weekly. This eventually paid off in the long run with respect to CO_2 emissions, fuel consumption, and cost.

The authors in [100] proposed a traffic-aware approach for offloading data from big-data-enabled ITS applications with a focus on discovering and selecting gateways. Their aim was to offer reliable communication with little delay for offloading data while alleviating network overhead incurred in gateway discovery. To achieve this aim, the authors designed an adaptive gateway advertisement algorithm that managed the frequency and area of advertisement dynamically using network and traffic reports in the gateway's environment. The proposed system was evaluated using practical simulation domain with respect to gateway access overhead and data offloading success ratio. Results showed a remarkable improvement in regards to permitting big traffic data centric ITS. Using New York City taxi and limousine commission datasets, the authors in [101] designed a system to analyze the runtime and predictive performance of several machine learning algorithms on a Spark cluster. They found that increasing the size of the dataset had an insignificant improvement on the accuracy of their classification, and the use of complex tree-ensemble methods contributed little to enhance the results produced by simpler algorithms.

5.3. Big Data Technology and Transportation

Big data describes blending advanced analytic techniques (such as machine learning and pattern recognition) together with a huge assembly of structured and unstructured, enormously and exponentially increasing in size, and extremely complex data to discover useful patterns, realize complex facts, trends, and relationships that exist in the data in order to improve decision making and process optimization [102]. Big data technologies refer to software frameworks that enable fast and efficient extraction, processing, and analysis of data from highly complex and vast datasets in a manner which conventional data management tools can never deal with. Big data has been employed in the transportation industry to resolve many conventional data-based challenges, and has enabled new applications, services, and opportunities [68].

5.3.1. Big Data Technological Frameworks

Some big data platforms have been used tremendously for research in SC transportation. This subsection reviews some big data technological frameworks for big data processing. It particularly describes some of the widely used technological platforms (i.e., software frameworks) for processing very large volumes of data and their use cases.

Apache Hadoop: Hadoop is a technology that satisfies the requirements of big data as it is scalable and designed as a software for big data processing. Hadoop is the most commonly used open source, distributed, and scalable software framework for storing and processing highly voluminous datasets. It offers a general big data process platform that enables the execution of diverse data analytical operations. Hadoop's capability to handle distributed processing makes it very suitable for analyzing data generated in ITS (e.g., smart card, social media, various sensors, GPS data, etc.). It was originally designed for use in the area of business intelligence, but its use has widened to various urban domains that utilize big data applications [104]. Hadoop enables the distribution of processing load among the nodes of the cluster which improves the processing strength. The number of nodes in a cluster can be increased or decreased depending on the need and can form homogenous clusters using diverse categories of machines to process unstructured data, rather than using a single costly supercomputer [105]. One main advantage of Hadoop is that it is inexpensive to use, since it is an open-source framework.

The Hadoop framework is made up of three major layers. The first layer is the Hadoop-distributed file system (HDFS) which uses the master–slave architecture to store, process, and analyze huge datasets to extricate relevant insight from data. The HDFS splits the big files and puts them into standard blocks and saves the blocks into the big cluster. In the second layer is the MapReduce, which has four basic elements (input, mapper, reducer, and output). Its operation involves splitting the big job into smaller jobs and executing them appropriately. At the third layer is the Yet Another Resource Negotiator (YARN), which is the resource manager responsible for allocating resource requests (CPU, memory) of the Hadoop cluster to different tasks. Because of Hadoop's popularity in big data processing in both industry and research communities, several of its extensions have been proposed such as HadoopGIS, HadoopTrajectory, HadoopDB, Hadoop++, etc. [106].

Apache Spark: Spark is a general and an extensively used distributed framework implemented using a large in-memory cluster of machines to better handle the processing of big data. Unlike Apache Hadoop, which is a disk-based system developed for I/O efficiency, Spark takes advantage of increased growth in the main memory capacity of machines that make up the cluster to attain better performance [107]. It permits the loading of user program data into the cluster and queries it repeatedly. Spark is very suitable for the deployment of machine learning methods and applications. The authors in [108] designed a method to analyze transportation data using Hadoop together with Spark to process transportation data in real-time. The architecture of the system is made up of four distinct layers—the data collection and acquisition layer, network layer, data processing layer, and application layer. The first employs methods that use sensors to collect data, and the collected data is then subjected to cleaning and transformation using data-filtering techniques to remove noise and unwanted information. At the processing layer, the data received from the data collection layer is further preprocessed and normalized using the Min-Max normalization technique. This technique was employed by the Hadoop ecosystem to efficiently process data [109], since the data was collected from diverse sources (e.g., road, parking). The proposed architecture provides modules for processing offline and online data. The Hadoop ecosystem was utilized to process offline data, while Spark together with the Hadoop ecosystem was used to process the online data. MapReduce of the Hadoop was used to deal with the big data. To optimize the processing power of the MapReduce, a scheduling algorithm was incorporated to split the task on the Hadoop server and to adaptively regulate the task on the Hadoop ecosystem using memory requirement and CPU utilization to move a task from one node to another.

Two drawbacks were observed in the Hadoop ecosystem—high performance nodes often switch to idle mode while low performance nodes maintain the active mode and remain busy always. To deal with these problems, the system incorporated a load balancer to monitor and balance the load among Hadoop and Spark nodes. After the data processing, the result (output) was stored using the HDFS. The HDFS enables the efficient and speedy processing of output data by the decision module. The proposed system incorporated HBASE in order to provide fast real-time access, and optimized caching of the output data. The output data from the Hadoop ecosystem is transferred to the application layer where it is used to make informed decisions. This work demonstrates how big data technologies enable the design and implementation of big data systems.

Python Library: Python recently has been utilized as a data analytics platform. The PyStack is well equipped with libraries that support data analysis, but a lot of these libraries are implemented to run on one CPU core and to handle data of similar sizes as that of the main memory. As a result, Python is not effective in processing big data. On the other hand, one can make use of PySpark to execute big data utilizing Spark on a cluster of machines. Presently, Spark remains the most commonly used distributed in-memory all-round execution engine that enables the use of a range of key libraries (e.g., SparkSQL, MLlib, etc.).

5.3.2. Big Data Architectures

This section reviews some of the big data architectures proposed by different researchers in this domain. The authors in [70] proposed a three-layered architecture of conducting big data analytics in ITS as shown in Figure 18. Their proposed architecture has three layers: (1) the data collection layer; (2) the data analytics layer; (3) the application layer. The data collection layer supplies the upper layers with the diverse data sources. The data analytics layer receives data from the previous data collection layer and then utilizes various machine learning and analytics techniques. The application layer utilizes the data processing results from the data analytics layer for various SC transportation scenarios (e.g., traffic signal control, traffic flow prediction, and traffic emergency rescue).

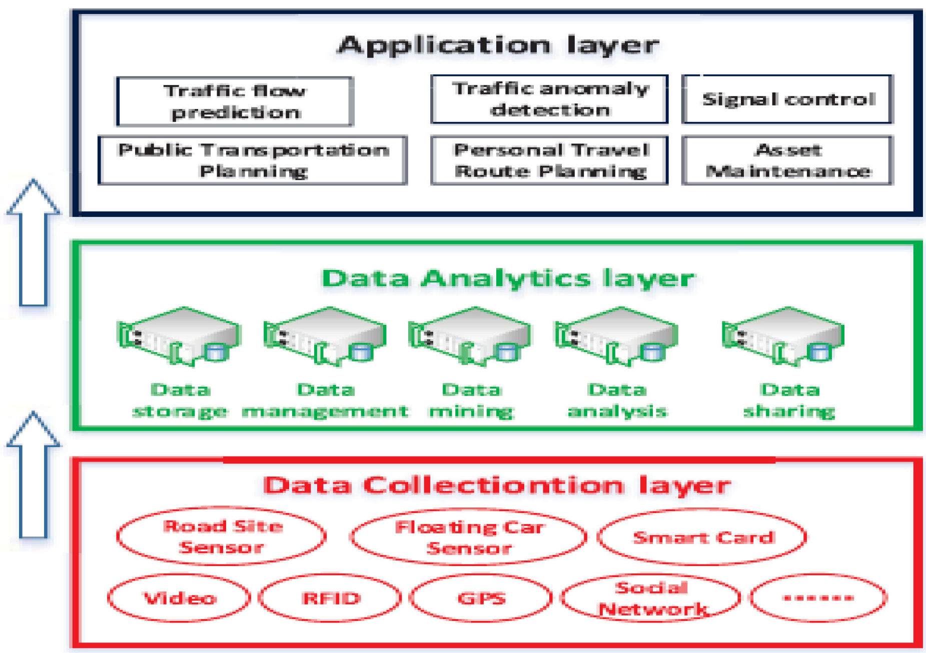

Figure 18. Three layered architectures for big data analytics [70].

The authors in [108] proposed a four-layered architecture to analyse big data and bring forward a system that can handle real-time transportation data. Figure 19 shows their proposed architecture. The architecture has four layers: (1) the data collection layer (2); the data communication layer; (3) the data processing layer; (4) the data application layer. Their proposed system collects data from various RFIDs installed at various locations in an SC. The RFIDs are programmed to identify malicious vehicles using their registration numbers, vehicle models, etc. The network layer consists of an SDN architecture to efficiently process and transport the data to the next layer. The data processing layer utilises various machine learning algorithms to efficiently analyse the data using the Hadoop platform. The application layer distributes the data to the users.

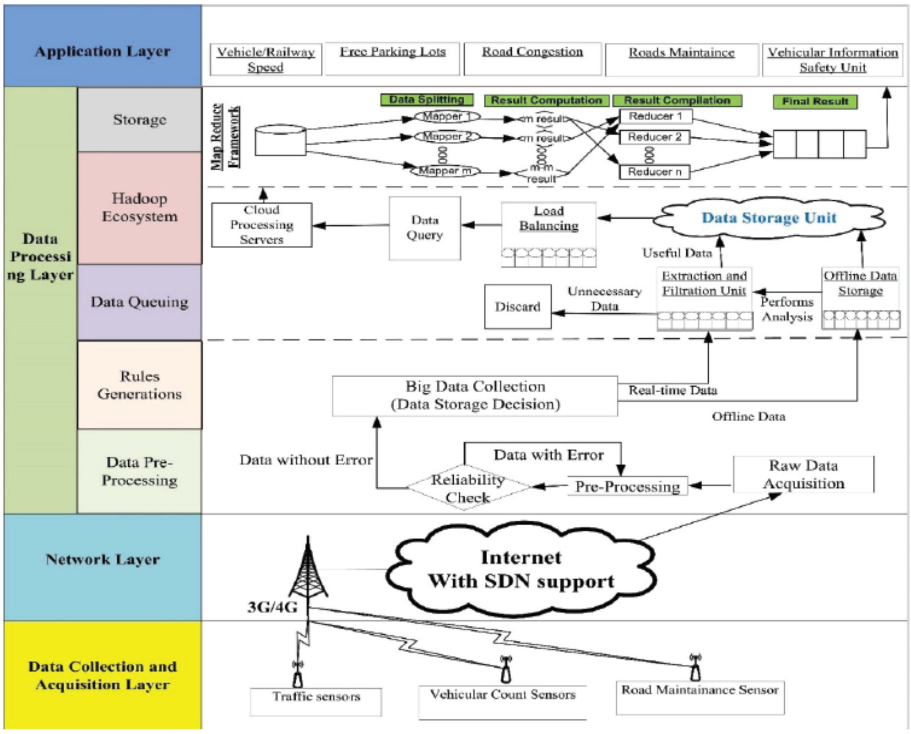

Figure 19. Four layered architectures for Big data analytics [110].

The authors in [111] proposed a big data analytics platform for smart urban transportation management (Figure 20). Their platform focuses on the City Administration Dashboard, which is a public transport analytics application that has been developed on top of the Europe-Brazil Collaboration of Big Data Scientific Research through Cloud-Centric Applications (EUBra-BIGSEA) platform. The City Administration Dashboard provides statistical trends about bus usage. The architecture includes several services such as PRIVAaaS (PRIVAcy as a Service), DQaaS (Data Quality as a Service), and EMaaS (Entity Matching as a Service).

5.3.3. Research Works on Big Data for Transportation

Researchers have utilized big data to try and solve challenges to improve the transportation system for SCs. Some of the problem areas that will be largely aided using big data includes transport planning, road safety, environment, traffic management, aviation, connected and autonomous vehicles (CAVs), freight and logistics, cost-effectiveness, railway, and data-related issues [102]. Major challenges related to planning includes: cross-media traffic operation, mobility on demand, changes in infrastructure to accommodate CAVs, ticketing, etc.

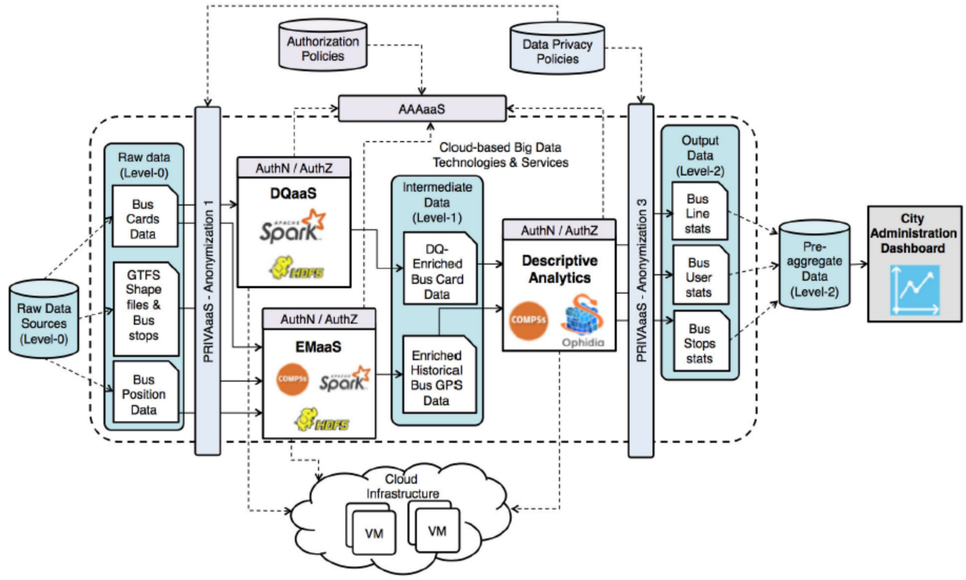

Figure 20. Integrated Big data analytics platform for smart urban transportation [111].

Using Shanghai traffic data alongside taxi and subway trajectory data, the authors in [113] designed a transit–transportation planning strategy between taxi and subways. Experiments showed that the strategy offered a very timely and bounded travel time using real city traffic, saving travel cost and time. Consequently, it can alleviate pressure on cities' road networks, minimize the general energy utilization of society, and expand the scope of public transport systems. The authors in [114] utilized vehicle positioning and smart card data to analyze the planning and operational processes of two real cases from Sweden and Netherlands. The processing and relevant insight obtained from data enabled good decision making and enhanced the public transport system.

To increase the efficient use of the road infrastructure, there is a need for dynamic traffic management. Big data applications have been utilized in major areas in traffic management, such as in congestion prevention and route planning, short-term forecasting, individualized travel information systems, parking request modelling, and fleet management, to provide insights at real-time. The authors in [115] proposed a graph-oriented scheme to analyze big data of traffic and vehicular networks, such as timestamps, geographical locations, and traffic intensity (speed and vehicle count) generated from IoT sensors installed on roads and vehicles to enable road users and traffic management authorities to make smart transport decisions. The architecture employed Graph tools along with parallel processing servers to realize real-time efficiency. The system implementation utilized Spark and Graph tools mounted on Hadoop parallel nodes to create and process graphs almost at real-time.

The authors in [116] designed a three-layer big data method for smart bus transportation management. The study was conducted with real world data acquired from roughly 2000 buses during a period of seven months from a vast bus network containing about 300 routes that span almost 5000 bus stops in the city of Fortaleza, Brazil. GPS tracking devices were installed on the buses to send the latitude and longitude coordinates every 15–30 s. Transaction data from automatic fare collection (AFC) machines at all bus stops were also collected. The big GPS signal and AFC data were processed using MapReduce to calculate the bus travel time and to determine travelers boarding location in the first layer. The authors in [117] discussed the role of big data analytics for smart transportation systems. The authors in [24] proposed a traffic flow prediction service based on cloud computing and big data analysis and conducted a traffic flow prediction based on the MapReduce framework, consisting of three stages: (1) model selection; (2) parameter estimation; (3) model combination. In [118], authors utilized big data generated from a microwave vehicle detection system (MVDS) installed on a motorway network in Orlando to design a dynamic strategy to monitor traffic and to simultaneously evaluate operation and safety in real-time. Their study confirmed that congestion significantly impacts the likelihood of a crash. The authors in [92] employed H2O and WEKA tools to analyze five classifiers on a big traffic accident dataset. Their analysis showed effectiveness in predicting accidents before they occurred and that traffic flow and safety decisions are tremendously impacted by drivers' behavior.

The authors in [120] designed a big data platform to show how big data can motivate large scale deployment of green vehicles. They used two datasets made up of 4.5 million trips and recorded parking events of 28,000 conventional fuel vehicles, which were monitored for more than one month period. The result of their analysis revealed the potential of big data for policy assessment towards low-emissions and the deployment of green vehicles. The authors in [121] conducted an analysis to investigate the driving range and factors influencing the energy consumption rate of completely battery-powered, electric vehicles (BEV) using real-world patterns and found that weather variables and driving pattern influence the driving range of BEVs in the real-world. The authors in [122] proposed a method to measure the carbon emission flow data of self-driving tour traffic from the year 2014 and evaluated its spatial connection with attractive locations based on data-mining techniques.

Rail transportation systems have enjoyed substantial transformation as a result of the increased attention and adoption of big data analytics in recent years [123]. This is because of their ability to generate and process huge amount of data that includes real-time train speed and location, departure and arrival time at certain stations, and passenger OD information. Big data analytics can enable operators of rail transport system to make informed decisions about train control and to improve operation efficiency. The areas where big data is expected to positively impact the rail transport system includes information management (e.g., passenger information, ticketing, tracking), train control (communication systems, automation, etc.), energy (intelligent power supply, smart metering etc.), infrastructure (track condition, signal systems, surveillance analytics etc.), and predictive maintenance (safety, real-time rescheduling). The authors in [124,125] proposed a big data analytic method that analyses rail big data and displays an overview of a selected railway network area. This enables operators to study operations and make informed decisions. The authors in [126] applied big data analytics to optimize the shortest path fare strategy of Beijing rail transit. Big data analytics was then used to process the basic data of passengers' entry and exit from stations to obtain a real travel time distribution from any origination to destination. The result of the analysis achieved by the proposed method will enable operators to determine the actual fare payable by passengers for any trip.

The authors in [127] proposed a model to analyze the cascading failure of weighted urban railway transit networks (URTNs). The model adopted the estimated operation interruption time of the station (EOITS) to distinguish the disturbance intensity of the system. The authors in [128] applied big data analytics to determine passengers' flow in Shenzhen's subway using smart card data. The result of the analysis enabled the derivation of policies for reducing urban congestion and optimizing network traffic. In [129], authors proposed a big data visualization framework to analyze passenger flow, while in [130], the authors proposed an approach for big data classification and transportation in rail networks. They presented a new framework for future data-driven railway condition monitoring systems (RCM). Figure 21 shows the proposed architecture for classified transportation of large-scale sensor data in rail networks. The architecture is composed of two main parts (data analysis and data transmission). The authors proposed two ways to perform classification on the data, which are signal analysis (SA)-based methods and machine learning (ML)-based approaches. To convert a signal from the time-domain to frequency-domain, discrete Fourier transformation (DFT) is used for time-frequency conversion. The one-class support vector machine (SVM) is selected for classification. The performance of the classifier was evaluated using real experimental data. Their work demonstrated good performances on various data sets under different operating conditions.

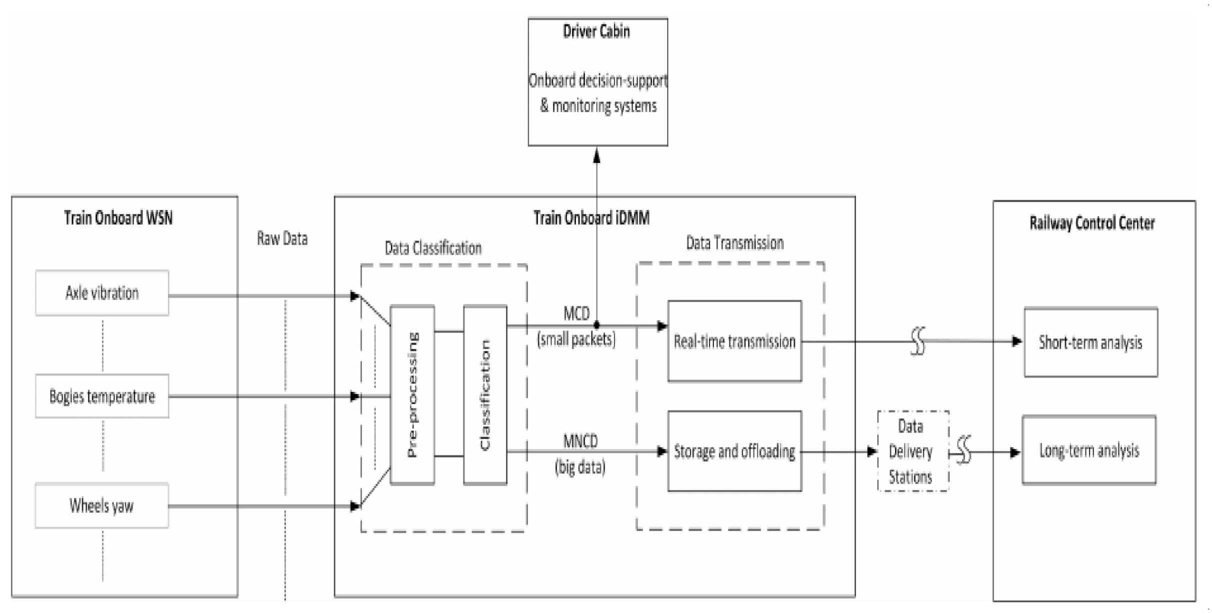

Figure 21. Big data classification and transportation in rail networks [130].

In achieving a smart aviation industry, big data is considered a key element. The authors in [102] categorized the challenges in the aviation sector into three types: (1) operations and air transport management (delay mitigation, weather resistance, flight plan optimization); (2) security and safety (on-board monitoring, privacy, air space operations in all weather, resilience against cyber-attacks); (3) incorporation of new technologies (deploying drones for air logistics). Big data has been employed to address some of these challenges. The authors in [132] proposed a hybrid analytic model to manage air cargo logistics. Their model combined a cluster and associative models for their analysis and incorporated the Diffusion of Innovative theory together with Resource Dependence theory into the analysis result to generate the operational strategy. The assessment of their strategy proved effective to improve cargo logistics.

The authors in [133] proposed a policy analytic framework for air connectivity in India and discussed how application of big data can help improve operational efficiency. A big data analytic method to optimize airline route profitability was presented by authors in [134], while the authors in [135] proposed a civil aircraft big data for civil aviation to facilitate developmental decisions. The authors in [136] proposed a big data-based security framework to preserve privacy and protect data.

6. Machine Learning Approaches for SC Transportation

Machine learning offers an auspicious avenue for SC transportation, considering its capability to exploit the power of data that has become increasingly available to SC transportation administrators and researchers. With the high volumes of data generated by SC transportation data sources (e.g., sensors, smart cards, videos, etc.) that cannot be examined individually, it is necessary to have a system that can learn and optimize on its own, based on previous experience. The operational changes in the context of SC transportation applications necessitate a generic, dynamic, and continuous learning technique. Therefore, for increased efficiency, it is critical to investigate the potentials of machine learning in the development of individualized services in SC transportation. Machine learning techniques are rapidly evolving due to improved algorithms, enhanced data collection methods, improved communication networks, new sensor/IO units, and interest in self-customization in response to user activity. The primary goal of machine learning is to effectively interpret new data and make predictions beyond the training sample, similar to real-time data. In general, machine-learning-based methods can offer descriptive (describe the current state of a system), predictive (predict the future state and values of a system) or prescriptive (recommend actions to maintain or improve system functionality) analyses. Researchers have proposed machine learning techniques and approaches to address real-world problems for SC transportation. The authors in [137] presented a study on intelligent transportation using machine learning. Their study explored machine learning in research and industry and focused on traffic management approaches for detection and prediction analyses. In the last decades, a considerable number of machine-learning-based studies have emerged in the literature, notably with a diverse use of multiple machine learning methods to investigate a variety of challenges in SC transportation. In this section, we discuss machine learning techniques for transportation applications in SCs. To minimize the volume of writing, some selected transportation applications have been presented in three critical areas—transportation and human mobility for SCs; traffic flow and density prediction; routing, planning, and route recommendation.

6.1. Transportation and Human Mobility for SCs

Human mobility data has been used for the analysis of a city and found to be useful for urban dynamics, planning and development. Current studies propose the extraction of the community structures of cities from human mobility data. The authors in [138] proposed an approach that utilizes network clustering methods using geographical cohesiveness and regularity from extracted clusters. Figure 22 shows the proposed network clustering for community structures. Their experimental results showed that the functional relations between city areas gave the best predictive information about the community structures of cities.

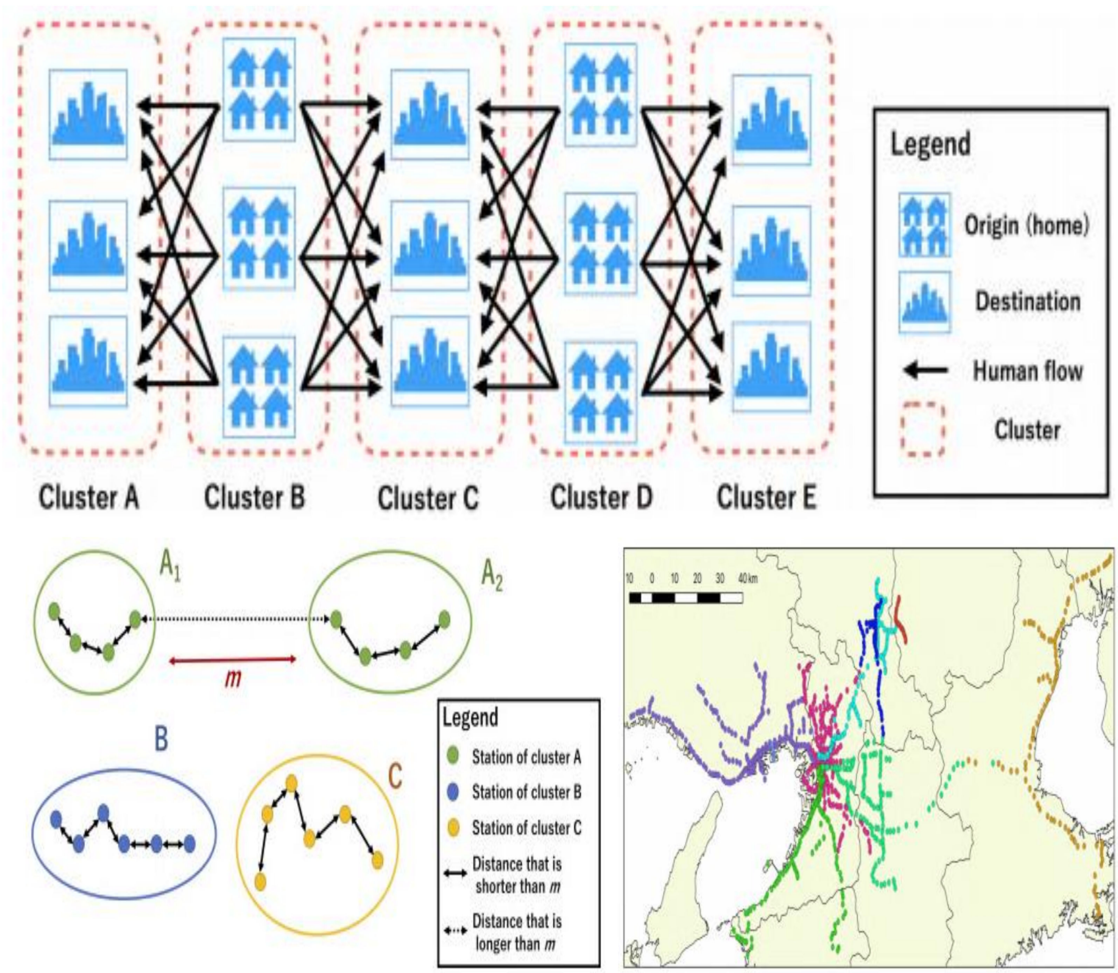

Figure 22. Network clustering for community structures [138].

6.2. Traffic Flow and Density Prediction

Various approaches have been proposed for traffic parameters' prediction. The authors in [139] proposed an approach to predict flight delay using a machine learning (ML) model and a convolution neural network (CNN) model on airline on-time performance (AOTP) and quality controlled local climatological datasets. The ML model achieved the best predictive result of 89.07% while the CNN model achieved a slightly better result, with 89.32% prediction accuracy. The authors in [140] proposed a model for short-term traffic volume prediction in transportation systems and presented a novel hybrid model (Gaussian process regression based on statistical learning theory and Bayesian theory) to predict the passenger flow volume. Their approach considered factors such as temporality, origin–destination spatiality, frequency and self-similarity, and historical probabilistic distribution perspectives. Their experimental results showed good performance even when the time intervals for traffic flow prediction were increased. In [141], authors proposed the online support vector machine for regression (OL-SVMR) approach to predict short-term traffic in typical and atypical conditions. Their experimental results showed good performance with the proposed approach. In [142], authors applied a framework that used the Haar cascade classifier and supervised learning (AdaBoost learning algorithms) to identify the directions of traffic streams and to extract the traffic flow parameters. The authors in [143] used GPS tracking systems in public transportation to analyze and predict passenger flow in real-time, while the authors in [144] proposed a short-term traffic flow prediction approach that utilizes wavelets and an extreme learning machine (ELM). An ERS-ELM (ensemble real-time sequential extreme learning machine) prediction approach was proposed in [145] for highway traffic peak and nonstationary states. The experimental results showed a high prediction accuracy of ERS-ELM, with an optimized training time.

The authors in [146] proposed an approach for the estimation of user demand in the public transportation network typified in the origin–destination matrix (ODM) from buses. The authors validated their model using data from the city of Quito. The authors in [147] proposed a hybrid forecasting model for short-term passenger flow prediction. Their approach utilized a combination of wavelet transformation (WT) and a kernel extreme learning machine (KELM). The authors validated their model using data from the city of Beijing. Their experimental results showed that the WT-KELM approach could give accurate information for the monitoring and early warning of urban rail transit. A study on machine learning algorithms for green, context-aware transportation systems was presented in [148]. The objective was to recommend the best transportation routes for the different means of transportation (train, metro, and bus) to reach a destination based on some user parameters. The authors in [149] performed a comparative analysis of four neural networks—two machine learning models based on back propagation neural networks (BPNN) and two deep learning models on recurrent neural networks (RNN). Their experimental results showed that models implemented on BPNN showed high performance when compared to those of RNN.

Besides, traditional approaches such as traffic sensors and devices, Internet data from social networks (e.g., Twitter) have become new sources for traffic flow prediction. The authors in [150] proposed a framework that retrieves and uses data from heterogeneous sources, including data from social networks, to detect traffic flows or patterns. Figure 23 shows the proposed framework. Their applications utilize data from various sources, such as entities extraction in tweets, event classifications, and classification of traffic states from image sources. Other research works on analytics from heterogeneous sources or multimodalities e.g., text, image, video, and speech, can be found in [151–153].

The authors in [154] proposed a neural network-based model for public transportation prediction using the traffic density matrix (as shown in Figure 24). The objective was to offer solutions to the bus arrival times at bus stations by considering local traffic conditions. The traffic conditions were represented in terms of a traffic density matrix. The network training was performed using stochastic gradient descent (SGD). In [155], authors proposed a real-time public transportation prediction with machine learning algorithms, including optimal least square (OLS) linear regression and support vector regression (SVR). Their work was validated on the SUMO (Simulation of Urban MObility) simulator [156]. Their experimental results showed that the proposed approach could outperform other approaches and reduce the mean absolute prediction error.

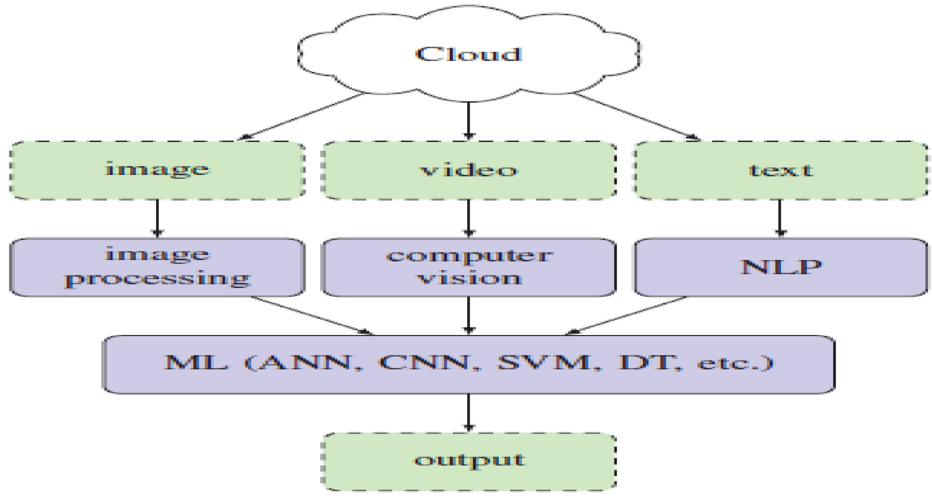

Figure 23. Framework from heterogeneous sources in intelligent transportation [150].

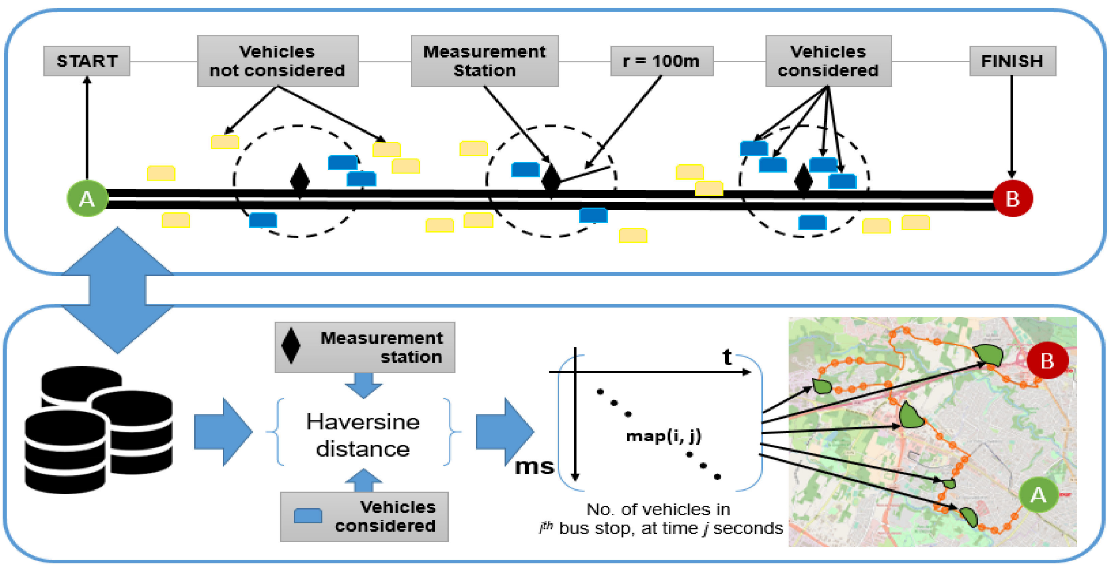

Figure 24. Neural network approach for public transportation prediction [154].

6.3. Routing, Planning and Route Recommendation

Machine learning can also be applied to assist users plan their travel trip and routes. To improve the utilization efficiency of public transportation services, the authors in [157] proposed a bus routing model that identifies and optimizes region pairs with flawed bus routes. The authors generated the human mobility patterns among regions using taxi traces and bus transactions. Their experimental results used real-world data collected in the city of Beijing, which contained 19 million taxi trips and 10 million bus trips. In [158], the authors proposed a hybrid solution for real-time travel mode detection and trip purpose prediction, which considered the use of a single preprocessing algorithm (using location traces obtained through smartphone sensors) for both problems. Experimental results showed accuracies of 88% for travel mode detection and 81% for trip purpose prediction. In [159], authors proposed a novel approach for public transport anonymous data collection.

7. Integrated Deep Learning towards SC Transportation

Deep learning models are inspired by the multi-layered structure of the human neural system. The authors in [160] surveyed the role of deep learning models in ITS and highlighted different types of deep learning models in the context of ITS for SC ecosystems. This section presents some representative studies for utilizing deep learning for SC transportation.

7.1. Routing and Planning

A Gaussian-prioritized approach to deploy additional routes for transportation was proposed in [161] using neural-networked-based passenger flow inference. Their experimental results recorded an improvement from 7 to 24% over the existing methods. In terms of route-based techniques, the authors in [162–164] studied a travel time prediction in both the forward and reverse trajectories based on deep belief networks (DBN), LSTM-based, and deep-travel models, respectively. The authors in [163] proposed an attention mechanism to aggregate context from local trajectory embeddings made by an LSTM (long short-term memory) network, while authors in [166] proposed a travel time estimation approach for the optimization of taxi–carpool systems. The authors in [167] proposed a multimodal transport recommender system using deep learning and tree models, which combined the weighted average ensemble method of CNN and GBDT (gradient-boosted decision trees). In order to enhance the categorical features for travel-mode preference prediction in deep learning, the authors in [168] proposed a deep neural network (DNN) architecture with entity embeddings (see Figure 25). The DNN architecture consists of four layers (embedding layer, concatenate layer, fully connected layer and output layer) and can efficiently learn the vector representations of the categorical data. The results of their experiments conducted on the London travel dataset showed that their approach with the entity-embedding technique outperformed other neural network models and tree-based models.

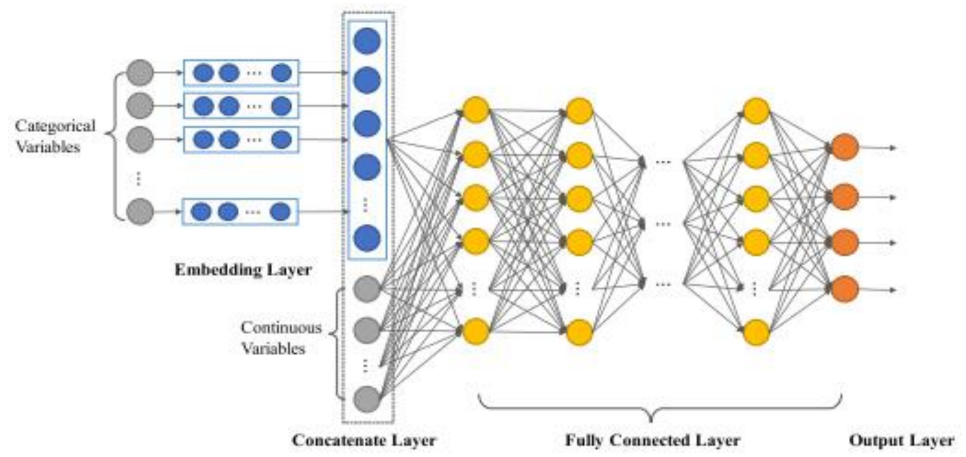

Figure 25. Deep neural network travel prediction framework [168].

7.2. Traffic Flow and Density Prediction

In recent years, traffic flow prediction has received extensive attention to prevent and minimize traffic congestions in SCs. Traffic prediction entails forecasting the traffic conditions, such as traffic volume and speed. The authors in [169] proposed a VGRAN (variational graph recurrent attention neural networks) model, utilizing a Bayesian framework for uncertainty-aware traffic forecasting, aimed to model the topology structure of road sensor networks and the spatial correlations among sensors to predict future traffic conditions. Figure 26 shows the architecture of the proposed VGRAN model. Their experimental work used two real-world traffic datasets: METR (dataset collected from Los Angeles County); and PEMS (dataset collected from Bay Area in California). The results showed that the proposed model outperformed other known state-of-the-art traffic speed forecasting models. The authors in [170] proposed an approach that models traffic flow over graph-like structures using a novel diffusion convolution layer.

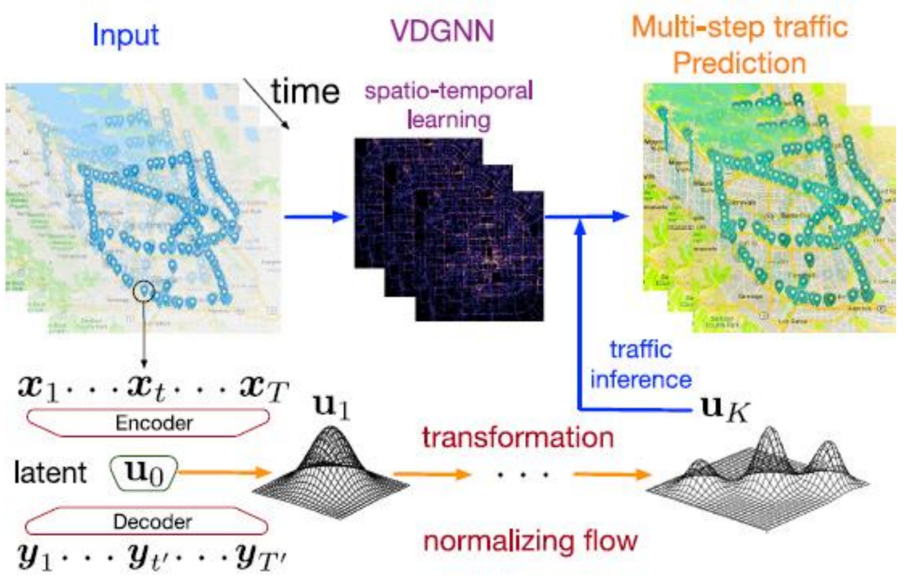

Figure 26. VGRAN framework for traffic forecasting [169].

The authors in [171] proposed a deep and embedding learning approach (DELA) that can learn from fine-grained traffic information, route structure, and weather conditions. Their proposed architecture consists of an embedding component (used to capture the categorical feature information and identify correlated features), a CNN component (used to learn the two-dimensional traffic flow data), and a LSTM component (used to maintain the memory of historical data). Figure 27 shows the analytical process, which has two stages: (1) data preprocess; (2) data analysis based on deep and embedding learning approaches. Their experimental results showed that the proposed approach could outperform existing methods in terms of prediction accuracy.

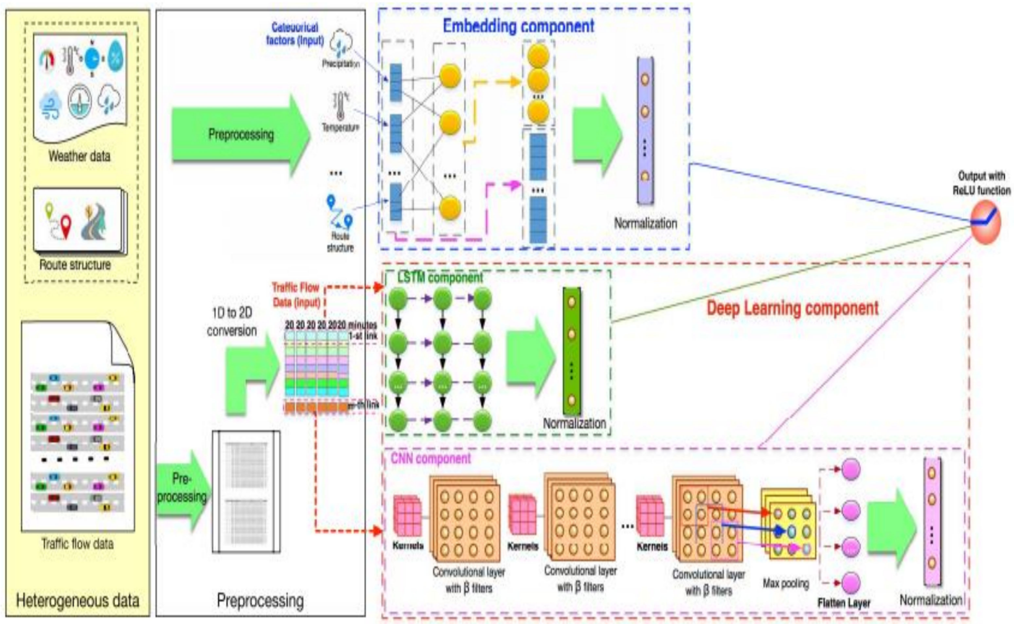

Figure 27. Deep learning approach for urban traffic flow prediction [171].

The authors in [172] considered the problem of predicting the in-flow and out-flow of traffic within a city and proposed a deep-learning-based approach, called ST-ResNet, to collectively forecast the in-flow and out-flow of crowds in each region of a city. A city is discretized to a two-dimensional grid and traffic flow is represented as a dense three-dimensional grid. Their work developed deep spatio-temporal residual networks for the prediction. The authors in [173] proposed a novel deep learning model termed as ST-3DNet (deep spatio-temporal 3D convolutional neural network) for traffic raster data prediction. This model uses 3D convolutions to capture the correlations of traffic data in the spatial and temporal dimensions. It consists of two major components to describe the two kinds of temporal properties of traffic data. ST-3DNet utilizes three-dimensional convolutions and blocks to model the two kinds of patterns and then aggregates them together in a weighted way for the final prediction. Their experimental results showed that ST-3DNet outperforms the state-of-the art baselines.

In [174], authors proposed a short-term traffic flow prediction model that combined the spatio-temporal analysis with a gated recurrent unit (GRU). A GRU is a type of recurrent neural network (RNN) and is a variant of the LSTM network. While maintaining the effect and making the structure simpler, it keeps the RNN prediction performance and has a significant increase in speed. GRU is simpler and has only two gates (update gate and the reset gate), while an LSTM network has three gate functions (input gate, forget gate, and output gate). Figure 28 shows the overall forecasting process. Their experiments compared the proposed model with the CNN model, and the results showed that the proposed method outperforms both in accuracy and stability.

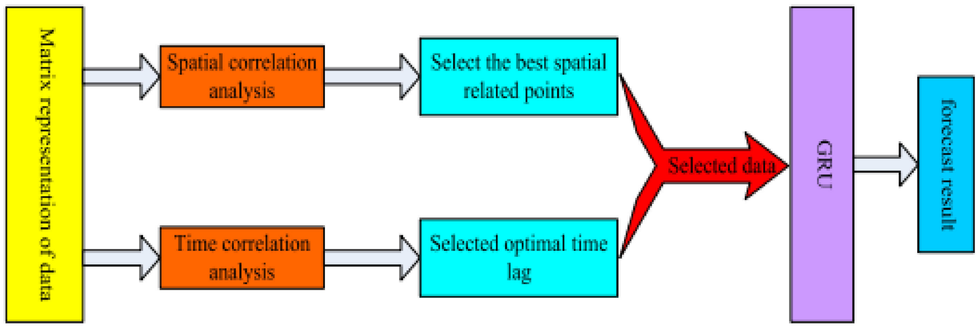

Figure 28. GRU framework for traffic prediction [174].

The authors in [175] proposed an ensemble model (EM) based on LSTM, DAE (deep autoencoder), and CNN for short-term traffic prediction. Their work considered the spatial and temporal characteristics of the traffic conditions. They evaluated their EM models on traffic data from two cities (California and London) and compared them with some well-known existing prediction models. Their experimental results showed that the EM can achieve better performance in terms of prediction accuracy.

Intelligent toll gates (ITGs) connect nearby metropolitan cities through smart highways; thus, they are also important infrastructure in SCs. Electronic toll collection (ETC) suffers from network limitations, such as optimal route utilization, long outstanding queues of connected smart vehicles (CSVs), fixed toll-pricing schemes for all CSVs, higher waiting time, variable delays, traffic congestion at toll gates, and complex payment mechanisms. The authors in [176] proposed DwaRa, which is a deep learning-based dynamic toll pricing scheme for ITS. The DwaRa system model is shown in Figure 29. In DwaRa, future traffic is predicted based on Markov queues to balance the congestion at different lanes at ITGs efficiently. An approach termed as the SI-LSTM (spatially induced-long-short-term memory) model is then used to predict current traffic and weather. For time series prediction of traffic which are updated in real-time, LSTM is a preferred choice.

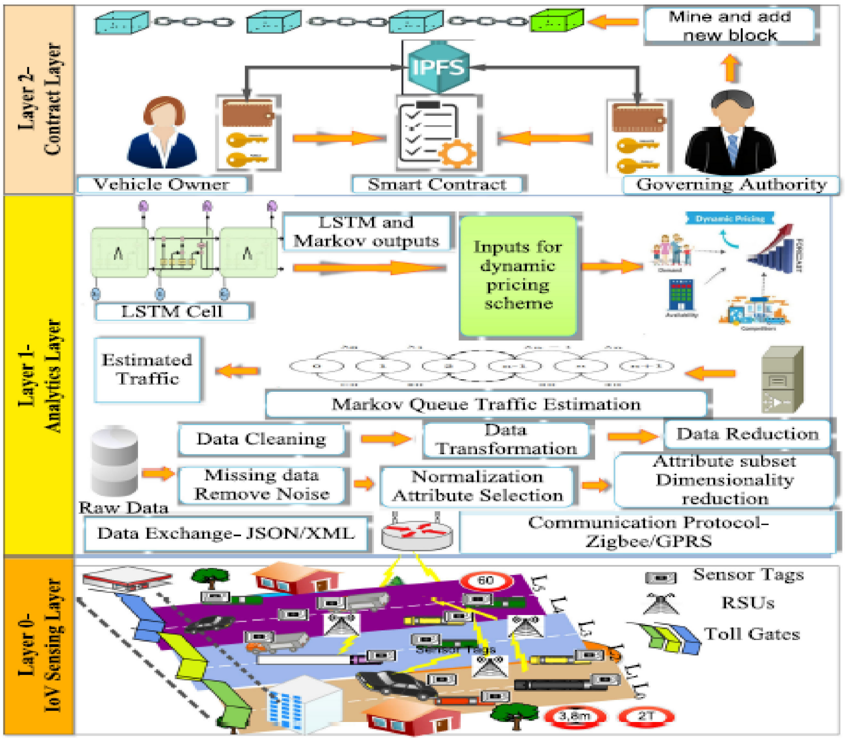

Figure 29. Dwara deep learning framework for dynamic toll pricing [176].

7.3. Passenger Flow in Public Transportation and City

The authors in [177] proposed a two-step K-means clustering model to capture passenger flow variation trends and ridership volume characteristics. They developed a predictability assessment model to recommend a reasonable time granularity interval to aggregate passenger flows. Then, a LSTM approach called CB-LSTM model, was proposed to conduct short-term passenger flow forecasting, based on the above clustering outcomes and the recommended time granularity interval. Figure 30 shows a framework of their model. Their experimental results showed that the proposed approach gave good predictive performance for short-term passenger flow on a network.

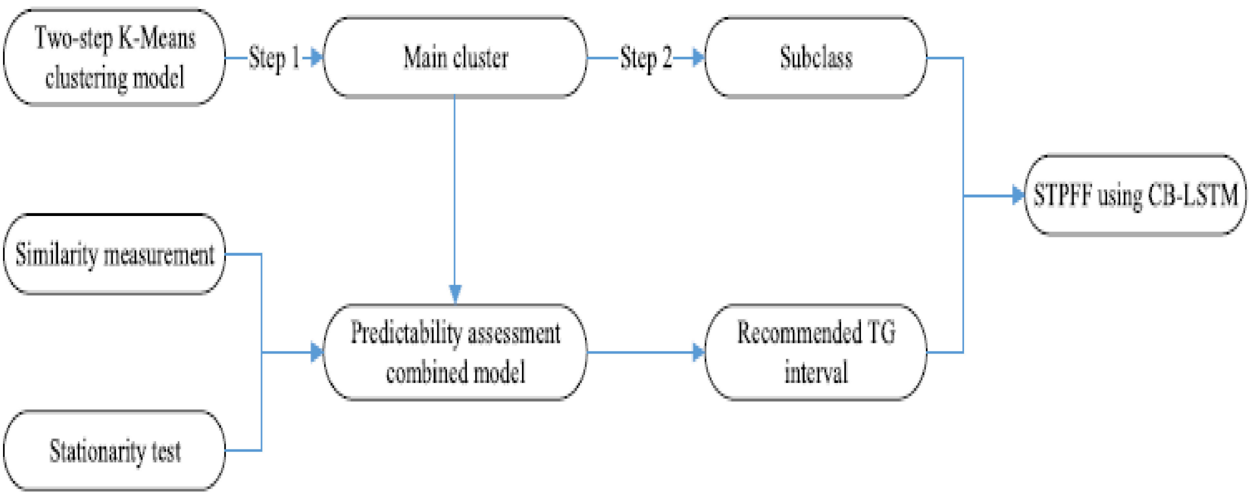

Figure 30. LSTM deep learning framework for passenger flow forecasting [177].

The authors in [178] combined deep learning (DL) and support vector machines (SVM) and proposed a DL-SVM model for urban rail transit (URT) passenger flow prediction. The deep belief network (DBN) was first used to extract the features and inherent variation of passenger flow data. Then SVM regression model was developed to predict passenger flow. Their experimental results showed that the DL-SVM outperforms the other models in accuracy and stability. Subway station passenger flow prediction is important for forecasting future passenger volume. The authors in [179] proposed a recurrent neural network RNN-based subway passenger flow rolling prediction. Their approach can help to inform safety warnings and evacuation passenger flow. The time series of passenger volumes was combined with weather data to create several supervised sequences, according to different values of timestep. Two artificial features were added as input to accelerate convergence. They used data from Shanghai traffic cards in their experiments. Their experimental results showed that the GRU network, with a timestep of 1.5 h gave the best performance for the long-term traffic flow rolling prediction. For short-term rolling prediction, GRU with a timestep of 45 min gave the best result.

In the rail transit system, due to the stochastic nature of the short-term dynamic passengers' origin and destination demand (OD matrix), accurate prediction of the distribution of passenger travel spatio-temporally is a challenge. The authors in [180] used an origin–destination matrix as input to a CNN-based model for predicting the in-flow and out-flow of nodes in a Beijing subway network. The authors in [181] proposed a combined multisource data with a deep learning method to improve the prediction of dynamic origin and destination demand (OD) matrix accuracy. The multisource data, such as smart card data, weather data, and mobile phone data were analyzed quantitatively based on the influencing factors; and 31 features were selected as model inputs. The authors in [182] also proposed a method that uses an LSTM to learn mobility patterns and to predict count-based traffic data between nodes in subway and bus networks.

8. Transportation Empowered by Artificial Intelligence (AI) and Other Techniques

AI is emerging as the driving force for new technologies and for the fast transition of intelligent transportation from mainly functional systems to truly intelligent and smart infrastructures. This section presents a review of the applications of AI techniques, such as evolutionary algorithms and swarm intelligence in the transportation industry. As the population of people in major cities continues to increase progressively, thus creating new mobility challenges, it is necessary for the transportation system to continue to evolve to effectively deal with the challenges. With the advent of ITS, which enables the massive generation of highly voluminous data, various AI techniques have been incorporated by the ITS to offer new services. These services normally require managing a notable volume of data generated by sensor devices, AFCs, GPS devices, social media, and smart phones [183].

Different AI techniques have been applied in several areas in ITS, such as traffic prediction and control, where systems are developed with the aim to minimize traffic congestion, control traffic signal, predict traffic volume, etc. The authors in [184] proposed a dynamic intelligent traffic light control system (DITLCS) based on fuzzy inference and deep reinforcement learning. The proposed DITLCS accepts real-time traffic information as input and makes adjustments to the traffic light duration dynamically. The proposed system operates in three modes (fair, priority, and emergency modes) where vehicles are categorized based on operational priority. The simulation results prove the efficiency of the proposed system when compared to recent algorithms on several performance parameters.

The authors in [185] proposed an intelligent traffic control system for vehicle passing at intersections using a back propagation (BP) neural network. Based on the model, the controller area network (CAN) communication network has been improved using the earliest deadline first (EDF) dynamic algorithm. The system was tested through simulations and proved very effective. The authors in [186] proposed an approach termed as BRBES (belief rule-based expert system) to control traffic signals at the intersection of roads. The proposal used belief rule base (BRB), which serves as the knowledge representation schema, while evidential reasoning serves as the inference engine. Their experimental results showed the proposed approach gave better reliability when compared to existing systems.

AI techniques have also been applied in the area of vehicle control systems to improve autonomous driving, reduce fuel and energy consumption, improve advanced braking systems, etc. The authors in [187] proposed an intelligent unidirectional and decentralized control method based on FLC for vehicle platooning. The controller performance was tuned by hybridizing the FLC using GA and proportional-integral-derivative (PID), and by the adaptation of FLC using neural networks to form the fuzzy x-tuned controllers to control the follower vehicles to attain their goals. Performance evaluation of each of each controller using simulations, regarding spacing error convergence and desired velocity tracking, revealed that all the controllers achieved their tasks amidst certain limitations.

The authors in [188] proposed a power control strategy using a genetic algorithm (GA) with a fuzzy logic controller (FLC) to efficiently control power transmission in electric vehicles. The simulation results exhibited that the proposed method was effective and superior. Using GA, the authors in [189] designed a simulation framework for charging the control system. The framework was used to simulate three scenarios and results showed that it can maximize the profit or reduce the charging time, depending on the objective of the various parking lots.

Urban transportation infrastructure affects immensely the quality of transport services. Implementing an optimal infrastructure enables the increased probability of considerably improving transport services [190]. The authors in [191] applied an evolutionary algorithm and a genetic operator to optimise the transit network design problem (TNDP). The authors noted that the increased use of private vehicles contributes immensely to traffic congestion, accidents, and pollution. They argued that having a viable, efficient and low-cost public bus network would discourage citizens from using personal vehicles. The simulation results of their proposal showed its effectiveness in addressing SC challenges. Using an evolving fuzzy neural network (EFNN), the authors in [192] predicted the travel speed multiple steps ahead using 2 min travel speed data obtained from remote traffic microwave sensors in the city of Beijing, and utilized a Takagi-Sugeno system to complement the fuzzy inference. The predictive performance of the proposed model was evaluated and compared against six traditional models. The EFNN produced better performance than those of the traditional models because of its strong learning ability. The authors in [193] aimed to improve the accuracy of traffic flow states by proposing a fusion clustering strategy for traffic flow state identification. The authors used three indices for evaluation (flow, velocity, and occupancy) for fusion. The optimized evaluation index weight was introduced into the clustering algorithm (fuzzy c-means) to realize traffic flow state identification based on multi-parameter fusion clustering. The process is shown in Figure 31.

The authors in [194] proposed and developed an innovative fuzzy logic approach to detect and predict the delay of public transport modes. The authors utilized predictive analytics and incorporated a knowledge variety of heterogeneous data, including transit data and weather data. The data was categorized based on fuzzy logic and random forest regression was applied to predict transit delays. The experimental works used transportation data from the city of Toronto. The authors in [195] proposed a hybrid RBF (radial basis function) neural network and fuzzy system for short-term road speed forecasting. Their work combined the fuzzy logic system with the RBF neural network. The authors in [196] proposed a fuzzy neural network model (FNM) to predict traffic flow in urban street networks. The authors in [184] proposed a fuzzy inference-enabled deep reinforcement learning-based traffic light control. The authors in [198] proposed a vehicle routing problem model with multiple fuzzy windows for the time-varying traffic flow. The performance of their model was confirmed through simulations and compared with the ant colony optimization (ACO) algorithms. Passenger flow is the basis for bus operation scheduling. The authors in [199] proposed the analysis of bus trip characteristic analysis and demanded forecasting based on GA-NARX (genetic algorithm optimized NARX neural network model). Their work combined the genetic algorithm with the neural network model. The GA-NARX was developed for the prediction to provide basic data for real-time scheduling and management of bus operations' management.

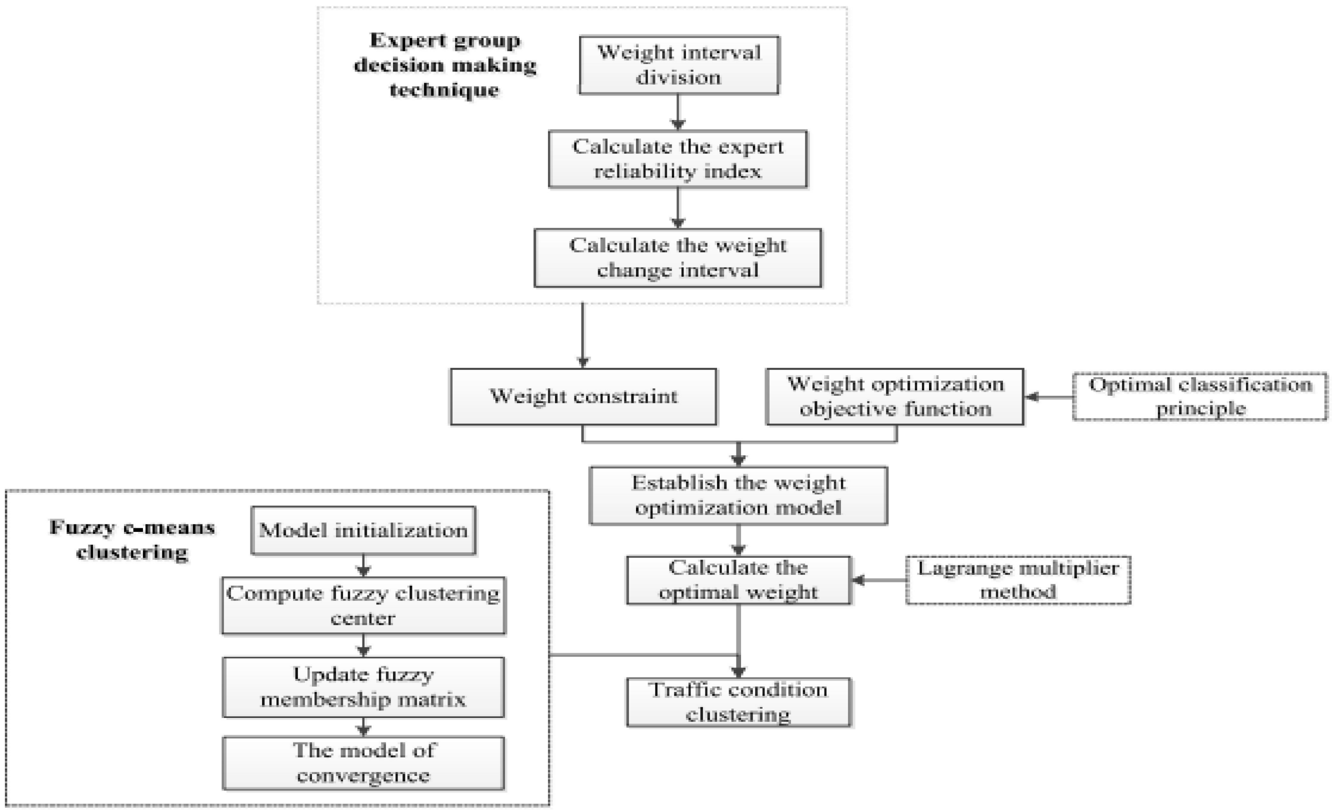

Figure 31. Traffic flow state identification based on multi-parameter fusion clustering [193].

There are some researchers who also worked on GA for smart transportation. The authors in [200] proposed a platform for public transportation management involving the optimal planning and scheduling of buses. Their approach considered the iterated local search (ILS) and genetic algorithm (GA), which are well-known methods in planning and scheduling. The authors in [201] proposed a traffic flow prediction model based on a wavelet neural network and achieved good prediction results. The identification of the characteristics of urban road traffic accidents is important. The authors in [202] proposed a feature recognition of urban road traffic accidents based on GA-XGBoost. The model was tested with data from traffic accidents in a sub-provincial city in China. Their experimental results showed a good predictive performance of the model to effectively identify the characteristics of urban road traffic accidents.

Particle swarm optimization (PSO) techniques simulate the social behaviors of nature such as bird flocking and fish schooling. Additionally, a multi-phase PSO algorithm is suitable to handle the combinatorial optimization problems of scheduling in railway freight transportation. To this effect, the authors in [203,204] proposed an improved multi-objective quantum-behaved PSO termed as IMOQPSO and IMOMPPSO (improved multi-objective multi-phase PSO) for the railway freight transportation routing design and their practical applications. Besides swarm intelligence, ant colony optimization (ACO) can also be used for route optimization. The authors in [205] proposed a route optimization for last-mile distribution of rural e-commerce logistics based on an improved ACO. Their experimental results showed that the improved ACO was effective on test datasets.

9. Conclusions

Owing to the large volumes of real-time data being generated daily as a result of the rapid surge in urban migration, the existing conventional data-processing tools are deficient to effectively realize the key targets of an SC transportation ecosystem. Consequently, this has brought enormous challenges for the SC transportation sector, including traffic congestion, fleet management/route planning problems, as well as the development of effective and efficient strategies to utilize the available infrastructures and to minimize traffic. To address these challenges, we have studied the state-of-the-art techniques for SC transportation applications with a focus on the emerging technologies from several information- and data-driven perspectives. This paper has given a comprehensive survey of the research area of smart transportation systems and emerging technologies. The paper contains core discussions on the impacts of geo-information on SC transportation, data-driven transportation, Big data technology, machine learning approaches for SC transportation, and recent trends using integrated deep learning towards SC transportation. The research findings in this survey paper give useful insights to researchers that demonstrate that data-driven approaches can be utilized for smart cities and transportation architecture. We also hope that this study will acquaint researchers with the recent trends and emerging technologies for SC transportation applications, as well as provides useful insights into how these technologies can be further exploited for SC transportation strategies. In essence, we have examined several use cases that can be exploited for SC transportation strategies.

The main limitation of this study is that it lacks some practical implementation of any of the use cases that were reviewed. Hence, our future work will present some practical considerations and implementation of any of the use cases. For instance, we intend to combine these approaches to implement a more efficient SC waste transport system.

Author Contributions: Conceptualization, Kenneth Li-Minn Ang and Jasmine Kah Phooi Seng; writing—original draft preparation, Jasmine Kah Phooi Seng, Ericmoore Ngharamike and Kenneth Li-Minn Ang; writing—review and editing, Gerald K. Ijemaru and Kenneth Li-Minn Ang. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

References

1. Batty, M.; Axhausen, K.W.; Giannotti, F.; Pozdnoukhov, A.; Bazzani, A.; Wachowicz, M.; Ouzounis, G.; Portugali, Y. Smart cities of the future. *Eur. Phys. J. Spéc. Top.* **2012**, *214*, 481–518. [CrossRef]
2. Ang, L.-M.; Seng, K.P.; Zungeru, A.M.; Ijemaru, G. Big Sensor Data Systems for Smart Cities. *IEEE Internet Things J.* **2017**, *4*, 1259–1271. [CrossRef]
3. Abdalla, R. *Introduction to Geospatial Information and Communication Technology (GeoICT)*; Springer: Singapore, 2016.
4. Weinberg, C. Is Alphabet Going to Build a City? *The Information*. 5 April 2016. Available online: https://www.theinformation.com/articles/is-alphabet-going-to-build-a-city (accessed on 30 November 2021).
5. Mohri, M.; Rostamizadeh, A.; Talwalkar, A. *Foundations of Machine Learning*; MIT Press: Cambridge, MA, USA, 2018.
6. Sun, H.; Liu, H.; Xiao, H.; He, R.R.; Ran, B. Use of Local Linear Regression Model for Short-Term Traffic Forecasting. *Transp. Res. Rec. J. Transp. Res. Board* **2003**, *1836*, 143–150. [CrossRef]
7. Vlahogianni, E.; Karlaftis, M.G.; Golias, J.C. Optimized and meta-optimized neural networks for short-term traffic flow prediction: A genetic approach. *Transp. Res. Part C Emerg. Technol.* **2005**, *13*, 211–234. [CrossRef]
8. Zenina, N.; Borisov, A. Regression Analysis for Transport Trip Generation Evaluation. *Inf. Technol. Manag. Sci.* **2013**, *16*, 89–94. [CrossRef]
9. Zhu, X.; Guo, J.; Huang, W.; Yu, F. Short Term Forecasting of Remaining Parking Spaces in Parking Guidance Systems. In Proceedings of the Transport Research Board Annual Meeting, Washington, DC, USA, 10–14 January 2016. No. 16-5060.
10. Bin, Y.; Zhongzhen, Y.; Baozhen, Y. Bus Arrival Time Prediction Using Support Vector Machines. *J. Intell. Transp. Syst.* **2006**, *10*, 151–158. [CrossRef]
11. Haydari, A.; Yilmaz, Y. Deep Reinforcement Learning for Intelligent Transportation Systems: A Survey. *IEEE Trans. Intell. Transp. Syst.* **2020**, *23*, 11–32. [CrossRef]
12. Chen, T. *Going Deeper with Convolutional Neural Network for Intelligent Transportation*; Worcester Polytechnic Institute: Worcester, MA, USA, 2015.
13. Duan, Y.; Lv, Y.; Kang, W.; Zhao, Y. A deep learning based approach for traffic data imputation. In Proceedings of the 17th International IEEE Conference on Intelligent Transportation Systems (ITSC), Qingdao, China, 8–11 October 2014; pp. 912–917.

14. Polson, N.G.; Sokolov, V.O. Deep learning for short-term traffic flow prediction. *Transp. Res. Part C Emerg. Technol.* **2017**, *79*, 1–17. [CrossRef]
15. Huang, W.; Hong, H.; Li, M.; Hu, W.; Song, G.; Xie, K. Deep architecture for traffic flow prediction. In Proceedings of the International Conference on Advanced Data Mining and Applications, Hangzhou, China, 14–16 December 2013; Springer: Berlin/Heidelberg, Germany, 2013; pp. 165–176.
16. Srivastava, S.K. Green supply-chain management: A state-of-the-art literature review. *Int. J. Manag. Rev.* **2007**, *9*, 53–80. [CrossRef]
17. Esmaeilian, B.; Wang, B.; Lewis, K.; Duarte, F.; Ratti, C.; Behdad, S. The future of waste management in smart and sustainable cities: A review and concept paper. *Waste Manag.* **2018**, *81*, 177–195. [CrossRef]
18. Govada, S.S.; Spruijt, W.; Rodgers, T. Smart city concept and framework. In *Smart Economy in Smart Cities*; Springer: Singapore, 2017; pp. 187–198.
19. Kumar, R.; Goel, S.; Sharma, V.; Garg, L.; Srinivasan, K.; Julka, N. A Multifaceted Vigilare System for Intelligent Transportation Services in Smart Cities. *IEEE Internet Things Mag.* **2020**, *3*, 76–80. [CrossRef]
20. Parmar, P.; Champaneria, T. Study and comparison of transportation system architectures for smart city. In Proceedings of the 2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), Palladam, India, 10–11 February 2017; pp. 675–680.
21. Ji, Z.; Ganchev, I.; O'Droma, M.; Zhang, X. A cloud-based intelligent car parking services for smart cities. In Proceedings of the 2014 31st URSI General Assembly and Scientific Symposium (URSI GASS), Beijing, China, 16–23 August 2014; pp. 1–4.
22. Vakula, D.; Raviteja, B. Smart public transport for smart cities. In Proceedings of the 2017 International Conference on Intelligent Sustainable Systems (ICISS), Thirupur, India, 12–13 December 2017; IEEE: New York, NY, USA, 2017. [CrossRef]
23. Lee, O.L.; Tay, R.I.; Too, S.T.; Gorod, A. A Smart City transportation System of Systems Governance Framework: A Case Study of Singapore. In Proceedings of the 2019 14th Annual Conference System of Systems Engineering (SoSE), Anchorage, AK, USA, 19–22 May 2019; pp. 37–42.
24. Zhu, F.; Li, Z.; Chen, S.; Xiong, G. Parallel Transportation Management and Control System and Its Applications in Building Smart Cities. *IEEE Trans. Intell. Transp. Syst.* **2016**, *17*, 1576–1585. [CrossRef]
25. Malik, F.; Shah, M.A.; Khattak, H.A. Intelligent Transport System: An Important Aspect of Emergency Management in Smart Cities. In Proceedings of the 2018 24th International Conference on Automation and Computing (ICAC), Newcastle upon Tyne, UK, 6–7 September 2018; pp. 1–6.
26. Lin, C.; Han, G.; Du, J.; Xu, T.; Shu, L.; Lv, Z. Spatiotemporal Congestion-Aware Path Planning Toward Intelligent Transportation Systems in Software-Defined Smart City IoT. *IEEE Internet Things J.* **2020**, *7*, 8012–8024. [CrossRef]
27. Chen, X.; Chen, R. A Review on Traffic Prediction Methods for Intelligent Transportation System in Smart Cities. In Proceedings of the 2019 12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), Suzhou, China, 19–21 October 2019; pp. 1–5.
28. Acunzo, D.; Zhu, Y.; Xie, B.; Baratoff, G. Context-Adaptive Approach for Vehicle Detection under Varying Lighting Conditions. In Proceedings of the 2007 IEEE Intelligent Transportation Systems Conference, Bellevue, WA, USA, 30 September–3 October 2007; pp. 654–660.
29. Farkas, K.; Fehér, G.; Benczúr, A.; Sidlo, C. Crowdsending based public transport information service in smart cities. *IEEE Commun. Mag.* **2015**, *53*, 158–165. [CrossRef]
30. Haklay, M.; Weber, P. OpenStreetMap: User-Generated Street Maps. *IEEE Pervasive Comput.* **2008**, *7*, 12–18. [CrossRef]
31. Cheng, G.; Guo, D.; Shi, J.; Qin, Y. Smart City-Wide Package Distribution Using Crowdsourced Public Transportation Systems. *IEEE Internet Things J.* **2019**, *6*, 7584–7594. [CrossRef]
32. Bergenhem, C.; Shladover, S.; Coelingh, E.; Englund, C.; Tsugawa, S. Overview of platooning systems. In Proceedings of the 19th ITS World Congress, Vienna, Austria, 22–26 October 2012.
33. Chen, C.; Zhang, Y.; Khosravi, M.R.; Pei, Q.; Wan, S. An Intelligent Platooning Algorithm for Sustainable Transportation Systems in Smart Cities. *IEEE Sens. J.* **2020**, *21*, 15437–15447. [CrossRef]
34. Menouar, H.; Guvenc, I.; Akkaya, K.; Uluagac, A.S.; Kadri, A.; Tuncer, A. UAV-Enabled Intelligent Transportation Systems for the Smart City: Applications and Challenges. *IEEE Commun. Mag.* **2017**, *55*, 22–28. [CrossRef]
35. Tripathy, A.K.; Tripathy, P.K.; Mohapatra, A.G.; Ray, N.K.; Mohanty, S.P. WeDoShare: A Ridesharing Framework in Transportation Cyber-Physical System for Sustainable Mobility in Smart Cities. *IEEE Consum. Electron. Mag.* **2020**, *9*, 41–48. [CrossRef]
36. Barth, M.; Todd, M. Intelligent transportation system architecture for a multi-station shared vehicle system. In Proceedings of the 2000 IEEE Intelligent Transportation Systems, Dearborn, MI, USA, 6 August 2002; pp. 240–245.
37. Mahmood, I.; Zubairi, J.A. Efficient Waste Transportation and Recycling: Enabling technologies for smart cities using the Internet of Things. *IEEE Electr. Mag.* **2019**, *7*, 33–43. [CrossRef]
38. Daniel, S.; Doran, M.A. geoSmartCity: Geomatics contribution to the smart city. In Proceedings of the 14th Annual International Conference on Digital Government Research, New York, NY, USA, 17–20 June 2013; pp. 65–71.
39. Doran, M.-A.; Daniel, S. Geomatics and Smart City: A transversal contribution to the Smart City development. *Inf. Polity* **2014**, *19*, 57–72. [CrossRef]
40. Roche, S. Geographic Information Science I: Why does a smart city need to be spatially enabled? *Prog. Hum. Geogr.* **2014**, *38*, 703–711. [CrossRef]

41. Nahiduzzaman, K.M.; Aldosary, A.S. City structure in transition: A conceptual discourse on the impact of information and communication technology (ICT). In *Social Development and High Technology Industries: Strategies and Applications*; IGI Global: Hershey, PA, USA, 2012; pp. 187–199.
42. Ahvenniemi, H.; Huovila, A.; Pinto-Seppä, I.; Airaksinen, M. What are the differences between sustainable and smart cities? *Cities* **2017**, *60*, 234–245. [CrossRef]
43. Aina, Y.A. Achieving smart sustainable cities with GeoICT support: The Saudi evolving smart cities. *Cities* **2017**, *71*, 49–58. [CrossRef]
44. Abukhater, A.; Walker, D. Making smart growth smarter with GeoDesign. *Directions Magazine*, 20 July 2010.
45. Liu, T.; Zhao, D.; Pan, M. An approach to 3D model fusion in GIS systems and its application in a future ECDIS. *Comput. Geosci.* **2016**, *89*, 12–20. [CrossRef]
46. Bremer, M.; Mayr, A.; Wichmann, V.; Schmidtner, K.; Rutzinger, M. A new multi-scale 3D-GIS-approach for the assessment and dissemination of solar income of digital city models. *Comput. Environ. Urban Syst.* **2016**, *57*, 144–154. [CrossRef]
47. Steiniger, S.; Neun, M.; Edwardes, A.; Lenz, B. Foundations of LBS, CartouCHe-Cartography Swiss Higher Education. *Obtido Em* **2008**, *20*, 2010.
48. Jiang, B.; Yao, X. Location-based services and GIS in perspective. *Comput. Environ. Urban Syst.* **2006**, *30*, 712–725. [CrossRef]
49. Batty, M. Big data, smart cities and city planning. *Dialog Hum. Geogr.* **2013**, *3*, 274–279. [CrossRef]
50. Chauhan, S.; Agarwal, N.; Kar, A.K. Addressing big data challenges in smart cities: A systematic literature review. *Info* **2016**, *18*, 73–90. [CrossRef]
51. Li, D.; Cao, J.; Yao, Y. Big data in smart cities. *Sci. China Inf. Sci.* **2015**, *58*, 1–12. [CrossRef]
52. Li, S.; Dragicevic, S.; Anton, F.; Sester, M.; Winter, S.; Çöltekin, A.; Pettit, C.; Jiang, B.; Haworth, J.; Stein, A.; et al. Geospatial big data handling theory and methods: A review and research challenges. *ISPRS J. Photogramm. Remote Sens.* **2016**, *115*, 119–133. [CrossRef]
53. Adnan, M.; Leak, A.; Longley, P. A geocomputational analysis of Twitter activity around different world cities. *Geo-Spat. Inf. Sci.* **2014**, *17*, 145–152. [CrossRef]
54. De Albuquerque, J.P.; Herfort, B.; Brenning, A.; Zipf, A. A geographic approach for combining social media and authoritative data towards identifying useful information for disaster management. *Int. J. Geogr. Inf. Sci.* **2015**, *29*, 667–689. [CrossRef]
55. Hawelka, B.; Sitko, I.; Beinat, E.; Sobolevsky, S.; Kazakopoulos, P.; Ratti, C. Geo-located Twitter as proxy for global mobility patterns. *Cartogr. Geogr. Inf. Sci.* **2014**, *41*, 260–271. [CrossRef] [PubMed]
56. Yang, W.; Mu, L.; Shen, Y. Effect of climate and seasonality on depressed mood among twitter users. *Appl. Geogr.* **2015**, *63*, 184–191. [CrossRef]
57. Aina, Y.A. Applications of geospatial technologies for practitioners: An emerging perspective of geospatial education. In *Emerging Informatics–Innovative Concepts and Applications*; InTech: Rijeka, Croatia, 2012; pp. 1–20.
58. Jackson, D.; Simpson, R. (Eds.) *D_City: Digital Earth | Virtual Nations | Data Cities: Connecting Global Futures for Environmental Planning*; DCity: Newton, NSW, Australia, 2013.
59. Pertence, A.A.; Mini, R.A.; Marques-Neto, H.T. Vulnerability Analysis of the Urban Transport System in the Context of Smart Cities. In Proceedings of the 2020 IEEE International Smart Cities Conference, Piscataway, NJ, USA, 28 September–1 October 2020; pp. 1–8.
60. Tong, L. Research on the Application of Key Technologies in the Construction of Smart Cities Based on Smart Transportation. In Proceedings of the 2020 IEEE Conference on Telecommunications, Optics and Computer Science (TOCS), Shenyang, China, 11–13 December 2020; pp. 189–194.
61. Htay, T.Z.; Komapov, V.G.; Glushenkov, V.A. Creation of an interactive geo-information model of the multifunctional magnetic levitation transport system ELTRO in Naypyidaw, Republic of Myanmar. In Proceedings of the 2021 3rd International Youth Conference on Radio Electronics, Electrical and Power Engineering (REEPE), Moscow, Russia, 12–14 March 2021; pp. 1–5.
62. Ghosh, K.; Musti, K.S. Integration of SLAM with GIS to model sustainable urban transportation system: A smart city perspective. In Proceedings of the 2020 12th International Conference on Computational Intelligence and Communication Networks (CICN), Bhimtal, India, 25–26 September 2020; pp. 261–267.
63. Timofeev, V.S.; Teselkina, K.V.; Veselova, A.S. Development and Research of Transport Speed Models Using the Methods of Geo-Statistical Data Analysis. In Proceedings of the 2018 XIV International Scientific-Technical Conference on Actual Problems of Electronics Instrument Engineering (APEIE), Novosibirsk, Russia, 2–6 October 2018; pp. 315–319.
64. Shanmugasundaram, R.; Santhiyakumari, N. Urban Sprawl Classification Analysis Using Image Processing Technique in Geoinformation System. In Proceedings of the 2018 Conference on Emerging Devices and Smart Systems (ICEDSS), Tiruchengode, India, 2–3 March 2018; pp. 192–196.
65. Chen, M.; Yu, X.; Liu, Y. TraLFM: Latent factor modeling of traffic trajectory data. *IEEE Trans. Intell. Transp. Syst.* **2019**, *20*, 4624–4634. [CrossRef]
66. Gokasar, I.; Cetinel, Y.; Baydogan, M.G. Estimation of Influence Distance of Bus Stops Using Bus GPS Data and Bus Stop Properties. *IEEE Trans. Intell. Transp. Syst.* **2019**, *20*, 4635–4642. [CrossRef]
67. Ghasemaghaei, M.; Calic, G. Assessing the impact of big data on firm innovation performance: Big data is not always better data. *J. Bus. Res.* **2020**, *108*, 147–162. [CrossRef]

68. Torre-Bastida, A.I.; Del Ser, J.; Laña, I.; Ilardia, M.; Bilbao, M.N.; Campos-Cordobés, S. Big Data for transportation and mobility: Recent advances, trends and challenges. *IET Intell. Transp. Syst.* **2018**, *12*, 742–755. [CrossRef]
69. Kaffash, S.; Nguyen, A.T.; Zhu, J. Big data algorithms and applications in intelligent transportation system: A review and bibliometric analysis. *Int. J. Prod. Econ.* **2021**, *231*, 107868. [CrossRef]
70. Zhu, L.; Yu, F.R.; Wang, Y.; Ning, B.; Tang, T. Big data analytics in intelligent transportation systems: A survey. *IEEE Trans. Intell. Transp. Syst.* **2019**, *20*, 383–398. [CrossRef]
71. Mohammed, S.; Arabnia, H.R.; Qu, X.; Zhang, D.; Kim, T.H.; Zhao, J. IEEE Access Special Section Editorial: Big Data Technology and Applications in Intelligent Transportation. *IEEE Access* **2020**, *8*, 201331–201344. [CrossRef]
72. Kim, K.; Gutesa, S.; Dimitrijevic, B.; Lee, J.; Spasovic, L.; Mirza, W.; Singh, J. Performance evaluation of video analytics for traffic incident detection and vehicle counts collection. In *Recent Advances in Intelligent Image Search and Video Retrieval*; Springer: Cham, Switzerland, 2017; pp. 213–231.
73. Wang, X.; Zhao, S.; Dong, L. Research and application of traffic visualization based on vehicle GPS big data. In Proceedings of the 2nd International Conference on Intelligent Transportation, Singapore, 25 November 2016; pp. 293–302.
74. Shukla, S.N.; Champaneria, T.A. Survey of various data collection ways for smart transportation domain of smart city. In Proceedings of the 2017 IEEE International Conference on i-Smac (Iot in Social, Mobile, Analytics and Cloud) (I-Smac), Palladam, India, 10–11 February 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 681–685.
75. Wen, S.; Gao, Y.; Zhang, D.; Yang, J.; Li, Q. An Efficient Data Acquisition System for Large Numbers of Various Vehicle Terminals. *IEEE Trans. Intell. Transp. Syst.* **2020**, 1–6. [CrossRef]
76. Chen, C.; Luan, T.H.; Guan, X.; Lu, N.; Liu, Y. Connected vehicular transportation: Data analytics and traffic-dependent networking. *IEEE Veh. Technol. Mag.* **2017**, *12*, 42–54. [CrossRef]
77. Lam, A.Y.; James, J.Q.; Hou, Y.; Li, V.O. Coordinated autonomous vehicle parking for vehicle-to-grid services: Formulation and distributed algorithm. *IEEE Trans. Smart Grid* **2018**, *9*, 4356–4366. [CrossRef]
78. Contreras-Castillo, J.; Zeadally, S.; Ibañez, J.A. Solving vehicular ad hoc network challenges with big data solutions. *IET Netw.* **2016**, *5*, 81–84. [CrossRef]
79. Zheng, X.; Chen, W.; Wang, P.; Shen, D.; Chen, S.; Wang, X.; Zhang, Q.; Yang, L. Big data for social transportation. *IEEE Trans. Intell. Transp. Syst.* **2015**, *17*, 620–630. [CrossRef]
80. Mounica, B.; Lavanya, K. Social Media data analysis for intelligent transportation systems. In Proceedings of the 2020 IEEE International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE), Vellore, India, 24–25 February 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 1–8.
81. Ghahramani, M.; Zhou, M.; Hon, C.T. Mobile Phone Data Analysis: A Spatial Exploration Toward Hotspot Detection. *IEEE Trans. Autom. Sci. Eng.* **2019**, *16*, 351–362. [CrossRef]
82. Cottrill, C.D.; Derrible, S. Leveraging Big Data for the Development of Transport Sustainability Indicators. *J. Urban Technol.* **2015**, *22*, 45–64. [CrossRef]
83. Zhu, L.; Yu, F.R.; Ning, B.; Tang, T. Cross-Layer Handoff Design in MIMO-Enabled WLANs for Communication-Based Train Control (CBTC) Systems. *IEEE J. Sel. Areas Commun.* **2012**, *30*, 719–728. [CrossRef]
84. Biuk-Aghai, R.P.; Kou, W.T.; Fong, S. Big data analytics for transportation: Problems and prospects for its application in China. In Proceedings of the 2016 IEEE Region 10 Symposium (TENSYMP), Bali, Indonesia, 9–11 May 2016; IEEE: Piscataway, NJ, USA, 2020; pp. 173–178.
85. Ayed, A.B.; Halima, M.B.; Alimi, A.M. Big data analytics for logistics and transportation. In Proceedings of the 2015 4th IEEE International Conference on Advanced Logistics and Transport (ICALT), Valenciennes, France, 20–22 May 2015; IEEE: Piscataway, NJ, USA, 2020; pp. 311–316.
86. Davenport, T.H.; Barth, P.; Bean, R. How'big data'is different. *MIT Sloan Manag. Rev.* **2012**, *54*, 43.
87. Wemegah, T.D.; Zhu, S. Big data challenges in transportation: A case study of traffic volume count from massive Radio Frequency Identification (RFID) data. In Proceedings of the 2017 IEEE International Conference on the Frontiers and Advances in Data Science (FADS), Xi'an, China, 23–25 October 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 58–63.
88. Ramesh, R.; Divya, G.; Dorairangaswamy, M.A.; Unnikrishnan, K.N.; Joseph, A.; Vijayakumar, A.; Mani, A. Real-Time Vehicular Traffic Analysis using Big Data Processing and IoT based Devices for Future Policy Predictions in Smart Transportation. In Proceedings of the 2019 IEEE International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 17–19 July 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1482–1488.
89. Syarif, M.; Adji, T.B. Big data analytics: Estimation of destination for users of bus rapid transit (BRT) public transportation in Jakarta. In Proceedings of the 2019 IEEE International Conference of Artificial Intelligence and Information Technology (ICAIIT), Yogyakarta, Indonesia, 13–15 March 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 209–214.
90. Lu, Y.; Misra, A.; Wu, H. Smartphone sensing meets transport data: A collaborative framework for transportation service analytics. *IEEE Trans. Mob. Comput.* **2017**, *17*, 945–960. [CrossRef]
91. Trueblood, F.; Rodriguez, D.; Hernandez, J.; Salomon, M.; Soundarajan, S.; Pirouz, M. Demystifying Transportation Using Big Data Analytics. In Proceedings of the 2019 IEEE International Conference on Computational Science and Computational Intelligence (CSCI), Las Vegas, NV, USA, 5–7 December 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1281–1286.

92. Al Najada, H.; Mahgoub, I. Anticipation and alert system of congestion and accidents in VANET using Big Data analysis for Intelligent Transportation Systems. In Proceedings of the 2016 IEEE Symposium Series on Computational Intelligence (SSCI), Athens, Greece, 6–9 December 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 1–8.
93. Lu, Y.; Wu, H.; Liu, X.; Chen, P. TourSense: A framework for tourist identification and analytics using transport data. *IEEE Trans. Knowl. Data Eng.* **2019**, *31*, 2407–2422. [CrossRef]
94. Puiu, D.; Bischof, S.; Serbanescu, B.; Nechifor, S.; Parreira, J.; Schreiner, H. A public transportation journey planner enabled by IoT data analytics. In Proceedings of the 2017 20th IEEE Conference on Innovations in Clouds, Internet and Networks (ICIN), Paris, France, 7–9 March 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 355–359.
95. Cyril, A.; George, V.; Mulangi, R.H. Electronic ticket machine data analytics for public bus transport planning. In Proceedings of the 2017 IEEE International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS), Chennai, India, 1–2 August 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 3917–3922.
96. Kwon, O.; Lee, N.; Shin, B. Data quality management, data usage experience and acquisition intention of big data analytics. *Int. J. Inf. Manag.* **2014**, *34*, 387–394.
97. Kong, X.; Li, M.; Tang, T.; Tian, K.; Moreira-Matias, L.; Xia, F. Shared subway shuttle bus route planning based on transport data analytics. *IEEE Trans. Autom. Sci. Eng.* **2018**, *15*, 1507–1520. [CrossRef]
98. Saroj, A.; Roy, S.; Guin, A.; Hunter, M.; Fujimoto, R. Smart city real-time data-driven transportation simulation. In Proceedings of the 2018 IEEE Winter Simulation Conference (WSC), Gothenburg, Sweden, 9–12 December 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 857–868.
99. Nelson, Q.; Steffensmeier, D.; Pawaskar, S. A simple approach for sustainable transportation systems in smart cities: A graph theory model. In Proceedings of the 2018 IEEE Conference on Technologies for Sustainability (SusTech), Long Beach, CA, USA, 11–13 November 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 1–5.
100. Darwish, T.S.; Bakar, K.A.; Kaiwartya, O.; Lloret, J. TRADING: Traffic Aware Data Offloading for Big Data Enabled Intelligent Transportation System. *IEEE Trans. Veh. Technol.* **2020**, *69*, 6869–6879. [CrossRef]
101. Howard, A.J.; Lee, T.; Mahar, S.; Intrevado, P.; Woodbridge, D.M.K. Distributed data analytics framework for smart transportation. In Proceedings of the 2018 20th IEEE International Conference on High Performance Computing and Communications; IEEE 16th International Conference on Smart City; IEEE 4th International Conference on Data Science and Systems (HPCC/SmartCity/DSS), Exeter, UK, 28–30 June 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 1374–1380.
102. Katrakazas, C.; Antoniou, C.; Vazquez, N.S.; Trochidis, I.; Arampatzis, S. Big data and emerging transportation challenges: Findings from the noesis project. In Proceedings of the 2019 6th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS), Cracow, Poland, 5–7 June 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1–9.
103. Demirkan, H.; Delen, D. Leveraging the capabilities of service-oriented decision support systems: Putting analytics and big data in cloud. *Decis. Support Syst.* **2013**, *55*, 412–421. [CrossRef]
104. Bibri, S.E. The IoT for smart sustainable cities of the future: An analytical framework for sensor-based big data applications for environmental sustainability. *Sustain. Cities Soc.* **2018**, *38*, 230–253. [CrossRef]
105. Singh, J.; Singla, V. Big data: Tools and technologies in big data. *Int. J. Comput. App.* **2015**, *112*, 6–10.
106. Alam, M.M.; Torgo, L.; Bifet, A. A Survey on Spatio-temporal Data Analytics Systems. *arXiv* **2021**, arXiv:2103.09883. [CrossRef]
107. Alam, M.M.; Ray, S.; Bhavsar, V.C. A Performance Study of Big Spatial Data Systems. In Proceedings of the 7th ACM SIGSPATIAL International Workshop on Analytics for Big Geospatial Data, Seattle, WA, USA, 6 November 2018; pp. 1–9.
108. Jan, B.; Farman, H.; Khan, M.; Talha, M.; Din, I.U. Designing a smart transportation system: An internet of things and big data approach. *IEEE Wirel. Commun.* **2019**, *26*, 73–79. [CrossRef]
109. Azeroual, O.; Fabre, R. Processing Big Data with Apache Hadoop in the Current Challenging Era of COVID-19. *Big Data Cogn. Comput.* **2021**, *5*, 12. [CrossRef]
110. Galliers, R.D.; Newell, S.; Shanks, G.; Topi, H. Datification and its human, organizational and societal effects. *J. Strateg. Inf. Syst.* **2017**, *26*, 185–190. [CrossRef]
111. Fiore, S.; Elia, D.; Pires, C.E.; Mestre, D.G.; Cappiello, C.; Vitali, M.; Andrade, N.; Braz, T.; Lezzi, D.; Moraes, R.; et al. An Integrated Big and Fast Data Analytics Platform for Smart Urban Transportation Management. *IEEE Access* **2019**, *7*, 117652–117677. [CrossRef]
112. Wang, Y.; Kung, L.; Byrd, T.A. Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. *Technol. Forecast. Soc. Change* **2018**, *126*, 3–13.
113. Guo, Y.; Wang, S.; Zheng, L.; Lu, M. Trajectory Data Driven Transit-Transportation Planning. In Proceedings of the 2017 5th IEEE International Conference on Advanced Cloud and Big Data (CBD), Shanghai, China, 13–16 August 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 380–384.
114. Van Oort, N.; Cats, O. Improving public transport decision making, planning and operations by using big data: Cases from Sweden and the Netherlands. In Proceedings of the 2015 18th IEEE international conference on intelligent transportation systems, Gran Canaria, Spain, 15–18 September 2015; IEEE: Piscataway, NJ, USA, 2015; pp. 19–24.
115. Rathore, M.M.; Ahmad, A.; Paul, A.; Jeon, G. Efficient graph-oriented smart transportation using internet of things generated big data. In Proceedings of the 2015 11th IEEE International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), Bangkok, Thailand, 23–27 November 2015; IEEE: Piscataway, NJ, USA, 2015; pp. 512–519.

116. Wang, Y.; Ram, S.; Currim, F.; Dantas, E.; Sabóia, L.A. A big data approach for smart transportation management on bus network. In Proceedings of the 2016 IEEE International Smart Cities Conference (ISC2), Trento, Italy, 12–15 September 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 1–6.
117. Shukla, S.; Balachandran, K.; Sumitha, V.S. A framework for smart transportation using Big Data. In Proceedings of the 2016 International Conference on ICT in Business Industry & Government, Indore, India, 18–19 November 2016; pp. 1–3.
118. Shi, Q.; Abdel-Aty, M. Big data applications in real-time traffic operation and safety monitoring and improvement on urban expressways. *Transp. Res. Part C Emerg. Technol.* **2015**, *58*, 380–394. [CrossRef]
119. Marshall, A.; Mueck, S.; Shockley, R. How leading organizations use big data and analytics to innovate. *Strategy Leadersh.* **2015**, *43*, 32–39. [CrossRef]
120. De Gennaro, M.; Paffumi, E.; Martini, G. Big data for supporting low-carbon road transport policies in Europe: Applications, challenges and opportunities. *Big Data Res.* **2016**, *6*, 11–25. [CrossRef]
121. Fetene, G.M.; Kaplan, S.; Mabit, S.L.; Jensen, A.F.; Prato, C.G. Harnessing big data for estimating the energy consumption and driving range of electric vehicles. *Transp. Res. Part D Transp. Environ.* **2017**, *54*, 1–11. [CrossRef]
122. Huang, Z.; Cao, F.; Jin, C.; Yu, Z.; Huang, R. Carbon emission flow from self-driving tours and its spatial relationship with scenic spots—A traffic-related big data method. *J. Clean. Prod.* **2017**, *142*, 946–955. [CrossRef]
123. Land, A.; Buus, A.; Platt, A. Data Analytics in Rail Transportation: Applications and Effects for Sustainability. *IEEE Eng. Manag. Rev.* **2020**, *48*, 85–91. [CrossRef]
124. Dimanche, V.; Goupil, A.; Philippot, A.; Riera, B.; Urban, A.; Gabriel, G. Massive Railway Operating Data Visualization; a Tool for RATP Operating Expert. *IFAC-PapersOnLine* **2017**, *50*, 15841–15846. [CrossRef]
125. Zhang, G.; Lin, J.; Zhang, Y.; Xue, K.; Nan, J.; Li, B. Big Data Based Intelligent Operation and Maintenance Platform. In Proceedings of the 2020 5th IEEE International Conference on Intelligent Transportation Engineering (ICITE), Beijing, China, 11–13 September 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 249–253.
126. Gao, H.; Liu, S.; Cao, G.; Zhao, P.; Zhang, J.; Zhang, P. Big Data Analysis of Beijing Urban Rail Transit Fares Based on Passenger Flow. *IEEE Access* **2020**, *8*, 80049–80062. [CrossRef]
127. Liu, B.; Zhu, G.; Li, X.; Sun, R. Vulnerability Assessment of the Urban Rail Transit Network Based on Travel Behavior Analysis. *IEEE Access* **2020**, *9*, 1407–1419. [CrossRef]
128. Wang, Y.; Zhang, W.; Zhang, F.; Yin, L.; Zhang, J.; Tian, C.; Jiang, W. Analysis of subway passenger flow based on smart card data. In Proceedings of the 2020 6th IEEE International Conference on Big Data Computing and Communications (BIGCOM), Deqing, China, 24–25 July 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 198–202.
129. Zhiyuan, H.; Liang, Z.; Ruihua, X.; Feng, Z. Application of big data visualization in passenger flow analysis of Shanghai Metro network. In Proceedings of the 2017 2nd IEEE International Conference on Intelligent Transportation Engineering (ICITE), Singapore, 1–3 September 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 184–188.
130. Saki, M.; Abolhasan, M.; Lipman, J. A novel approach for big data classification and transportation in rail networks. *IEEE Trans. Intell. Transp. Syst.* **2019**, *21*, 1239–1249. [CrossRef]
131. Saboo, A.R.; Kumar, V.; Park, I. Using big data to model time-varying effects for marketing resource (re) allocation. *Mis Q.* **2016**, *40*, 911–939. [CrossRef]
132. Wu, P.J.; Yang, C.K. Unstructured big data analytics for air cargo logistics management. In Proceedings of the 2018 IEEE International Conference on Service Operations and Logistics, and Informatics (SOLI), Singapore, 31 July–2 August 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 274–278.
133. Sankaranarayanan, H.B.; Thind, R.S. Multi-modal travel in India: A big data approach for policy analytics. In Proceedings of the 2017 7th IEEE International Conference on Cloud Computing, Data Science & Engineering-Confluence, Noida, India, 12–13 January 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 243–248.
134. Kasturi, E.; Devi, S.P.; Kiran, S.V.; Manivannan, S. Airline Route profitability analysis and Optimization using BIG DATA analyticson aviation data sets under heuristic techniques. *Procedia Comput. Sci.* **2016**, *87*, 86–92. [CrossRef]
135. Li, S.; Yang, Y.; Yang, L.; Su, H.; Zhang, G.; Wang, J. Civil aircraft big data platform. In Proceedings of the 2017 11th IEEE International Conference on Semantic Computing (ICSC), San Diego, CA, USA, 30 January–1 February 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 328–333.
136. Zhijun, W.; Caiyun, W. Security-as-a-service in big data of civil aviation. In Proceedings of the 2015 IEEE International Conference on Computer and Communications (ICCC), Chengdu, China, 10–11 October 2015; pp. 240–244.
137. Alsrehin, N.O.; Klaib, A.F.; Magableh, A. Intelligent transportation and control systems using data mining and machine learning techniques: A comprehensive study. *IEEE Access* **2019**, *7*, 49830–49857. [CrossRef]
138. Maeda, T.N.; Mori, J.; Hayashi, I.; Sakimoto, T.; Sakata, I. Comparative examination of network clustering methods for extracting community structures of a city from public transportation smart card data. *IEEE Access* **2019**, *7*, 53377–53391. [CrossRef]
139. Jiang, Y.; Liu, Y.; Liu, D.; Song, H. Applying Machine Learning to Aviation Big Data for Flight Delay Prediction. In Proceedings of the 2020 IEEE International Conference on Dependable, Autonomic and Secure Computing, International Conference on Pervasive Intelligence and Computing, International Conference on Cloud and Big Data Computing, International Conference on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech), Calgary, AB, Canada, 17–22 August 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 665–672.

140. Diao, Z.; Zhang, D.; Wang, X.; Xie, K.; He, S.; Lu, X.; Li, Y. A hybrid model for short-term traffic volume prediction in massive transportation systems. *IEEE Trans. Intell. Transp. Syst.* **2018**, *20*, 935–946. [CrossRef]
141. Castro-Neto, M.; Jeong, Y.S.; Jeong, M.K.; Han, L.D. Online-SVR for short-term traffic flow prediction under typical and atypical traffic conditions. *Exp. Syst. Appl.* **2009**, *36*, 6164–6173. [CrossRef]
142. Ke, R.; Li, Z.; Tang, J.; Pan, Z.; Wang, Y. Real-time traffic flow parameter estimation from UAV video based on ensemble classifier and optical flow. *IEEE Trans. Intell. Transp. Syst.* **2018**, *20*, 54–64. [CrossRef]
143. Zhang, J.; Shen, D.; Tu, L.; Zhang, F.; Xu, C.; Wang, Y.; Tian, C.; Li, X.; Huang, B.; Li, Z. A real-time passenger flow estimation and prediction method for urban bus transit systems. *IEEE Trans. Intell. Transp. Syst.* **2017**, *18*, 3168–3178. [CrossRef]
144. Feng, W.; Chen, H.; Zhang, Z. Short-term traffic flow prediction based on wavelet function and extreme learning machine. In Proceedings of the 2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS), Beijing, China, 24–26 November 2017; pp. 531–535.
145. Wang, D.; Xiong, J.; Xiao, Z.; Li, X. Short-term traffic flow prediction based on ensemble real-time sequential extreme learning machine under non-stationary condition. In Proceedings of the 2016 IEEE 83rd Vehicular Technology Conference (VTC Spring), Nanjing, China, 15–18 May 2016; pp. 1–5.
146. Zapata, L.P.; Flores, M.; Larios, V.; Maciel, R.; Antúnez, E.A. Estimation of people flow in public transportation network through the origin-destination problem for the South-Eastern corridor of Quito city in the smart cities context. In Proceedings of the 2019 IEEE International Smart Cities Conference (ISC2), Casablanca, Morocco, 14–17 October 2019; pp. 181–186.
147. Liu, R.; Wang, Y.; Zhou, H.; Qian, Z. Short-term passenger flow prediction based on wavelet transform and kernel extreme learning machine. *IEEE Access* **2019**, *7*, 158025–158034. [CrossRef]
148. Said, A.M.; Abd-Elrahman, E.; Afifi, H. A comparative study on machine learning algorithms for green context-aware intelligent transportation systems. In Proceedings of the 2017 International Conference on Electrical and Computing Technologies and Applications (ICECTA), Ras AI Khaimah, United Arab Emirates, 21–23 November 2017; pp. 1–5.
149. Heghedus, C.; Chakravorty, A.; Rong, C. Neural network frameworks. Com$parison on public transportation prediction. In Proceedings of the 2019 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW), Rio de Janeiro, Brazil, 20–24 May 2019; pp. 842–849.
150. Bazzan, A.L.; Chamby-Diaz, J.C.; Estevam, R.S.; Schmidt, L.D.A.; Pasin, M.; Samatelo, J.L.A.; Ribeiro, M.V.L. Using Information from Heterogeneous Sources and Machine Learning in Intelligent Transportation Systems. In Proceedings of the 2019 IEEE 15th International Conference on Intelligent Computer Communication and Processing (ICCP), Cluj-Napoca, Romania, 5–7 September 2019; pp. 213–220.
151. Seng, K.P.; Ang, L.M. A big data layered architecture and functional units for the multimedia Internet of Things. *IEEE Trans. Multi-Scale Comput. Syst.* **2018**, *4*, 500–512. [CrossRef]
152. Seng, J.K.P.; Ang, K.L.M. Multimodal Emotion and Sentiment Modeling From Unstructured Big Data: Challenges, Architecture, & Techniques. *IEEE Access* **2019**, *7*, 90982–90998.
153. Shoumy, N.J.; Ang, L.M.; Seng, K.P.; Rahaman, D.M.; Zia, T. Multimodal big data affective analytics: A comprehensive survey using text, audio, visual and physiological signals. *J. Netw. Comput. Appl.* **2019**, *149*, 102447. [CrossRef]
154. Panovski, D.; Scurtu, V.; Zaharia, T. A neural network-based approach for public transportation prediction with traffic density matrix. In Proceedings of the 2018 7th European Workshop on Visual Information Processing (EUVIP), Tampere, Finland, 26–28 November 2018; pp. 1–6.
155. Panovski, D.; Zaharia, T. Real-time public transportation prediction with machine learning algorithms. In Proceedings of the 2020 IEEE International Conference on Consumer Electronics (ICCE), Las Vegas, NV, USA, 4–6 January 2020; pp. 1–4.
156. Panovski, D.; Zaharia, T. Simulation-based vehicular traffic lights optimization. In Proceedings of the 2016 12th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), Naples, Italy, 28 November–1 December 2016; pp. 258–265.
157. Liu, Y.; Liu, C.; Yuan, N.J.; Duan, L.; Fu, Y.; Xiong, H.; Xu, S.; Wu, J. Intelligent bus routing with heterogeneous human mobility patterns. *Knowl. Inf. Syst.* **2017**, *50*, 383–415. [CrossRef]
158. Soares, E.F.D.S.; Revoredo, K.; Baião, F.; de MS Quintella, C.A.; Campos, C.A.V. A combined solution for real-time travel mode detection and trip purpose prediction. *IEEE Trans. Intell. Transp. Syst.* **2019**, *20*, 4655–4664. [CrossRef]
159. Minea, M.; Dumitrescu, C.; Chiva, I.C. Unconventional public transport anonymous data collection employing artificial intelligence. In Proceedings of the 2019 11th International Conference on Electronics, Computers and Artificial Intelligence (ECAI), Pitesti, Romania, 27–29 June 2019; pp. 1–6.
160. Veres, M.; Moussa, M. Deep learning for intelligent transportation systems: A survey of emerging trends. *IEEE Trans. Intell. Transp. Syst.* **2019**, *21*, 3152–3168. [CrossRef]
161. Lin, F.; Fang, J.Y.; Hsieh, H.P. A Gaussian-Prioritized Approach for Deploying Additional Route on Existing Mass Transportation with Neural-Network-Based Passenger Flow Inference. In Proceedings of the 2020 IEEE Congress on Evolutionary Computation (CEC), Glasgow, UK, 19–24 July 2020; pp. 1–8.
162. Siripanpornchana, C.; Panichpapiboon, S.; Chaovalit, P. Travel-time prediction with deep learning. In Proceedings of the 2016 IEEE Region 10 Conference (TENCON), Singapore, 22–25 November 2016; pp. 1859–1862.
163. Wang, D.; Zhang, J.; Cao, W.; Li, J.; Zheng, Y. When will you arrive? Estimating travel time based on deep neural networks. In Proceedings of the AAAI Conference on Artificial Intelligence, New Orleans, LA, USA, 2–7 February 2018.

164. Zhang, H.; Wu, H.; Sun, W.; Zheng, B. Deeptravel: A neural network based travel time estimation model with auxiliary supervision. *arXiv* **2018**, arXiv:1802.02147.
165. Zhu, X.; Li, J.; Liu, Z.; Wang, S.; Yang, F. Learning transportation annotated mobility profiles from GPS data for context-aware mobile services. In Proceedings of the 2016 IEEE International Conference on Services Computing (SCC), San Francisco, CA, USA, 27 June–2 July 2016; pp. 475–482.
166. Jindal, I.; Qin, Z.T.; Chen, X.; Nokleby, M.; Ye, J. Optimizing taxi carpool policies via reinforcement learning and spatio-temporal mining. In Proceedings of the 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 10–13 December 2018; pp. 1417–1426.
167. Abedalla, A.; Fadel, A.; Tuffaha, I.; Al-Omari, H.; Omari, M.; Abdullah, M.; Al-Ayyoub, M. MTRecS-DLT: Multi-modal transport recommender system using deep learning and tree models. In Proceedings of the 2019 Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS), Granada, Spain, 22–25 October 2019; pp. 274–278.
168. Ma, Y.; Zhang, Z. Travel Mode Choice Prediction Using Deep Neural Networks with Entity Embeddings. *IEEE Access* **2020**, *8*, 64959–64970. [CrossRef]
169. Zhou, F.; Yang, Q.; Zhong, T.; Chen, D.; Zhang, N. Variational Graph Neural Networks for Road Traffic Prediction in Intelligent Transportation Systems. *IEEE Trans. Ind. Inform.* **2020**, *17*, 2802–2812. [CrossRef]
170. Li, Y.; Yu, R.; Shahabi, C.; Liu, Y. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. *arXiv* **2017**, arXiv:1707.01926.
171. Zheng, Z.; Yang, Y.; Liu, J.; Dai, H.N.; Zhang, Y. Deep and embedded learning approach for traffic flow prediction in urban informatics. *IEEE Trans. Intell. Transp. Syst.* **2019**, *20*, 3927–3939. [CrossRef]
172. Zhang, J.; Zheng, Y.; Qi, D. Deep spatio-temporal residual networks for citywide crowd flows prediction. In Proceedings of the AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 4–9 February 2017.
173. Guo, S.; Lin, Y.; Li, S.; Chen, Z.; Wan, H. Deep spatial–temporal 3D convolutional neural networks for traffic data forecasting. *IEEE Trans. Intell. Transp. Syst.* **2019**, *20*, 3913–3926. [CrossRef]
174. Dai, G.; Ma, C.; Xu, X. Short-term traffic flow prediction method for urban road sections based on space–time analysis and GRU. *IEEE Access* **2019**, *7*, 143025–143035. [CrossRef]
175. Zheng, G.; Chai, W.K.; Katos, V. An Ensemble Model for Short-Term Traffic Prediction in Smart City Transportation System. In Proceedings of the 2019 IEEE Global Communications Conference (GLOBECOM), Waikoloa, HI, USA, 9–13 December 2019; pp. 1–6.
176. Shukla, A.; Bhattacharya, P.; Tanwar, S.; Kumar, N.; Guizani, M. Dwara: A deep learning-based dynamic toll pricing scheme for intelligent transportation systems. *IEEE Trans. Veh. Technol.* **2020**, *69*, 12510–12520. [CrossRef]
177. Zhang, J.; Chen, F.; Shen, Q. Cluster-based LSTM network for short-term passenger flow forecasting in urban rail transit. *IEEE Access* **2019**, *7*, 147653–147671. [CrossRef]
178. Zhu, K.; Xun, P.; Li, W.; Li, Z.; Zhou, R. Prediction of passenger flow in urban rail transit based on big data analysis and deep learning. *IEEE Access* **2019**, *7*, 142272–142279. [CrossRef]
179. Sha, S.; Li, J.; Zhang, K.; Yang, Z.; Wei, Z.; Li, X.; Zhu, X. RNN-based subway passenger flow rolling prediction. *IEEE Access* **2020**, *8*, 15232–15240. [CrossRef]
180. Li, J.; Peng, H.; Liu, L.; Xiong, G.; Du, B.; Ma, H.; Wang, L.; Bhuiyan, M.Z.A. Graph CNNs for urban traffic passenger flows prediction. In Proceedings of the 2018 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI), Guangzhou, China, 8–12 October 2018; pp. 29–36.
181. Li, D.; Cao, J.; Li, R.; Wu, L. A Spatio-Temporal Structured LSTM Model for Short-Term Prediction of Origin-Destination Matrix in Rail Transit with Multisource Data. *IEEE Access* **2020**, *8*, 84000–84019. [CrossRef]
182. Toqué, F.; Côme, E.; El Mahrsi, M.K.; Oukhellou, L. Forecasting dynamic public transport origin-destination matrices with long-short term memory recurrent neural networks. In Proceedings of the 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC), Rio de Janeiro, Brazil, 1–4 November 2016; pp. 1071–1076.
183. Machin, M.; Sanguesa, J.A.; Garrido, P.; Martinez, F.J. On the use of artificial intelligence techniques in intelligent transportation systems. In Proceedings of the 2018 IEEE Wireless Communications and Networking Conference Workshops (WCNCW), Barcelona, Spain, 15–18 April 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 332–337.
184. Kumar, N.; Rahman, S.S.; Dhakad, N. Fuzzy inference enabled deep reinforcement learning-based traffic light control for intelligent transportation system. *IEEE Trans. Intell. Transp. Syst.* **2020**, *22*, 4919–4928. [CrossRef]
185. Lv, Z.; Lou, R.; Singh, A.K. AI Empowered Communication Systems for Intelligent Transportation Systems. *IEEE Trans. Intell. Transp. Syst.* **2021**, *22*, 4579–4587. [CrossRef]
186. Hossain, M.S.; Sinha, H.; Mustafa, R. A belief rule based expert system to control traffic signals under uncertainty. In Proceedings of the 2015 International Conference on Computer and Information Engineering (ICCIE), Rajshahi, Bangladesh, 26–27 November 2015; IEEE: Piscataway, NJ, USA, 2015; pp. 83–86.
187. Olwan, M.A.; Mostafa, A.A.; AbdELAty, Y.M.; Mahfouz, D.M.; Shehata, O.M.; Morgan, E.I. Behavior Evaluation of Vehicle Platoon via Different Fuzzy-X Tuned Controllers. In Proceedings of the 2020 8th EEE International Conference on Control, Mechatronics and Automation (ICCMA), Moscow, Russia, 6–8 November 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 149–155.

188. Wang, W.C.; Tai, C.C.; Wu, S.J.; Liu, Z.Y. A hybrid genetic algorithm with fuzzy logic controller for wireless power transmission system of electric vehicles. In Proceedings of the 2015 IEEE International Conference on Industrial Technology (ICIT), Seville, Spain, 17–19 March 2015; IEEE: Piscataway, NJ, USA, 2015; pp. 2622–2627.
189. Yeh, Y.C.; Tsai, M.S. Development of a genetic algorithm based electric vehicle charging coordination on distribution networks. In Proceedings of the 2015 IEEE Congress on Evolutionary Computation (CEC), Sendai, Japan, 25–28 May 2015; IEEE: Piscataway, NJ, USA, 2015; pp. 283–290.
190. Alaguvelu, R.; Curry, D.M.; Dagli, C.H. Fuzzy—Genetic algorithm approach to generate an optimal meta-architecture for a smart, safe & efficient city transportation system of systems. In Proceedings of the 2016 11th IEEE System of Systems Engineering Conference (SoSE), Kongsberg, Norway, 12–16 June 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 1–6.
191. Nayeem, M.A.; Islam, M.M.; Yao, X. Solving transit network design problem using many-objective evolutionary approach. *IEEE Trans. Intell. Transp. Syst.* **2019**, *20*, 3952–3963. [CrossRef]
192. Tang, J.; Liu, F.; Zou, Y.; Zhang, W.; Wang, Y. An improved fuzzy neural network for traffic speed prediction considering periodic characteristic. *IEEE Trans. Intell. Transp. Syst.* **2017**, *18*, 2340–2350. [CrossRef]
193. Bao, D. A multi-index fusion clustering strategy for traffic flow state identification. *IEEE Access* **2019**, *7*, 166404–166409. [CrossRef]
194. Leung, C.K.; Elias, J.D.; Minuk, S.M.; de Jesus, A.R.R.; Cuzzocrea, A. An Innovative Fuzzy Logic-Based Machine Learning Algorithm for Supporting Predictive Analytics on Big Transportation Data. In Proceedings of the 2020 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), Glasgow, UK, 19–24 July 2020; pp. 1–8.
195. Ai, C.; Jia, L.; Hong, M.; Zhang, C. Short-term road speed forecasting based on hybrid RBF neural network with the aid of fuzzy system-based techniques in urban traffic flow. *IEEE Access* **2020**, *8*, 69461–69470. [CrossRef]
196. Yin, H.; Wong, S.; Xu, J.; Wong, C.K. Urban traffic flow prediction using a fuzzy-neural approach. *Transp. Res. Part C Emerg. Technol.* **2002**, *10*, 85–98. [CrossRef]
197. Schmidhuber, J. Deep learning in neural networks: An overview. *Neural Netw.* **2015**, *61*, 85–117. [CrossRef] [PubMed]
198. Zheng, J. A Vehicle Routing Problem Model With Multiple Fuzzy Windows Based on Time-Varying Traffic Flow. *IEEE Access* **2020**, *8*, 39439–39444. [CrossRef]
199. Sun, F.; Wang, X.L.; Zhang, Y.; Liu, W.X.; Zhang, R.J. Analysis of bus trip characteristic analysis and demand forecasting based on GA-NARX neural network model. *IEEE Access* **2020**, *8*, 8812–8820. [CrossRef]
200. Thiranjaya, C.; Rushan, R.; Udayanga, P.; Kaushalya, U.; Rankothge, W. Towards a Smart City: Application of Optimization for a Smart Transportation Management System. In Proceedings of the 2018 IEEE International Conference on Information and Automation for Sustainability (ICIAfS), Colombo, Sri Lanka, 21–22 December 2018; pp. 1–6.
201. Yang, H.J.; Hu, X. Wavelet neural network with improved genetic algorithm for traffic flow time series prediction. *Optik* **2016**, *127*, 8103–8110. [CrossRef]
202. Qu, Y.; Lin, Z.; Li, H.; Zhang, X. Feature recognition of urban road traffic accidents based on GA-XGBoost in the context of big data. *IEEE Access* **2019**, *7*, 170106–170115. [CrossRef]
203. Zhang, Q.; Liu, S.; Gong, D.; Zhang, H.; Tu, Q. An improved multi-objective quantum-behaved particle swarm optimization for railway freight transportation routing design. *IEEE Access* **2019**, *7*, 157353–157362. [CrossRef]
204. Wang, R.; Zhang, Z.; Wang, L.; Zhang, H. Resource-Expandable Railway Freight Transportation Routing Optimization. *IEEE Access* **2019**, *7*, 161607–161620. [CrossRef]
205. Liu, W. Route optimization for last-mile distribution of rural E-commerce logistics based on ant colony optimization. *IEEE Access* **2020**, *8*, 12179–12187. [CrossRef]