applied sciences

MDPI

Article

Smart Mobility with Big Data: Approaches, Applications, and Challenges

Dohoon Lee [1], David Camacho [2] and Jason J. Jung [1,*]

1 Department of Computer Engineering, Chung-Ang University, 84, Heukseok-ro, Dongjak-gu, Seoul 06974, Republic of Korea; dhlee7@cau.ac.kr
2 Departamento de Sistemas Informáticos, Universidad Politécnica de Madrid, 28031 Madrid, Spain; david.camacho@upm.es
* Correspondence: j3ung@cau.ac.kr

Abstract: Many vehicles are connected to the Internet, and big data are continually created. Various studies have been conducted involving the development of artificial intelligence, machine learning technology, and big data frameworks. The analysis of smart mobility big data is essential and helps to address problems that arise as society faces increased future mobility. In this paper, we analyze application issues such as personal information leakage and data visualization due to increased data exchange in detail, as well as approaches focusing on analyses exploiting machine learning and architecture research exploiting big data frameworks, such as Apache Hadoop, Apache Spark, and Apache Kafka. Finally, future research directions and open challenges are discussed.

Keywords: big data; smart mobility; transportation; big data analysis

check for updates

Citation: Lee, D.; Camacho, D.; Jung, J.J. Smart Mobility with Big Data: Approaches, Applications, and Challenges. *Appl. Sci.* **2023**, *13*, 7244. https://doi.org/10.3390/app13127244

Academic Editor: Wenjie Zhang

Received: 5 April 2023
Revised: 1 June 2023
Accepted: 8 June 2023
Published: 17 June 2023

1. Introduction

In the era of the fourth Industrial Revolution, the importance and value of big data are emphasized. Thus, numerous forms of transportation data are collected for analysis. Exploiting transportation data is critical. Accidents and congestion can be predicted using data analysis, and security issues are also prominent with the advent of connected vehicles, which exchange data among themselves. Thus, research on smart mobility and transportation big data has been conducted, such as regression analysis using transportation data, the investigation of security issues between vehicle networks, and prediction of accidents and congestion using machine learning and deep learning. Smart mobility is a combination of information technology and personal transportation, such as shared electric kickboards, autonomous vehicles, and unmanned aerial vehicles. In this paper, we introduce analysis methods and applications for smart mobility and explore future research directions.

Smart mobility is a practical example of the Internet of Things (IoT), whereby mobile users can communicate over a wireless network. Accordingly, users can make the best choice for traffic routes while increasing safety, minimizing travel time, reducing congestion and environmental pollution, and maximizing productivity. In addition, the Canadian global market research firm Precedence Research expects the global smart mobility market to grow to USD 57.88 billion by 2022 and to USD 250.3 billion by 2030 [1]. Thus, research on smart mobility and big data analysis is very valuable and significant.

This paper is organized into six sections. Section 2 describes the paper's research method and citation statistics, and Section 3 references papers that explain smart mobility and big data analysis methods. Section 4 analyzes transportation big data and describes applications applied to smart mobility. In Section 5, we suggest future research trends and the limitations of research. Finally, in Section 6, we present our conclusions suggest future research directions.

The contributions of this paper are as follows:

- A comprehensive review of approaches to transportation big data, such as analysis and data processing methods proposed in recent years;
- An introduction to smart mobility services and applications of the latest trends, such as artificial intelligence, big data frameworks, machine learning, and security;
- Exploration of the limitations of the current study and directions for future research.

2. Bibliographic Analysis

This paper closely analyzes 96 papers—94 published between 2015 and 2023, one in 2012, and one in 2014—with more than 30 citations and an average of 110 citations. China was the country that published the most papers in this domain, with 21 submitted papers; followed by the United States of America, with 15; Republic of Korea, with 14 published papers. The number of submissions by country was determined by the nationality of the institution, not by the nationalities of the authors. In addition, 6 out of 92 papers were published by IEEE Transactions on Intelligent Transportation Systems, the most common journal. Figure 1 presents the most common keywords for the reviewed papers in a word cloud.

Figure 1. Keywords of the analyzed papers in a word cloud.

3. Methods and Technologies for Smart Mobility

In this section, we discuss how to analyze and store smart mobility data and exploit machine learning algorithms and big data frameworks. We also provide a detailed review of approaches, citing various studies on efficient storage and analysis of big datasets that grow over time.

3.1. Machine Learning

In this section, we detail the results of smart mobility big data analysis using various machine learning algorithms. Machine learning is a representative method for the implementation of artificial intelligence and is a subset that focuses on building systems that learn or improve performance based on data. It has a narrower scope than artificial intelligence and a more comprehensive meaning than deep learning.

3.1.1. Predicting Traffic Accidents

Radar sensors and rear cameras have been used to prevent or reduce accidents. However, research on accident prevention and anomaly detection of vehicles has not yet reached the commercialization stage due to other aspects of the road surface. In this section, we review research on accident prevention methods using smart mobility.

In [2], the authors proposed SafeDrive to detect anomalous driving behavior. SafeDrive detects abnormal driving habits through unsupervised learning according to features such as rapid acceleration, sudden braking, RPM–speed mismatch, speeding, RPM anomalies, rapid swerving, and neutral taxing. In [3], the Hadoop framework was used to efficiently process and analyze numerous traffic data using sampling methods to solve the data imbalance problems. Accordingly, the prediction system first preprocesses and analyzes big data on traffic to generate data for the learning system. The imbalance in the generated data is corrected using the sampling method. The modified data were classified into several groups, and classification analysis was applied to improve the prediction accuracy. In addition, a MapReduce-based synthetic minority oversampling technique (SMOTE) was devised to solve the data imbalance.

In [4], the authors predicted the exact arrival time using linear regression, naive Bayes, and distributed random forest to reflect it in the algorithm before accidents and congestion occur in order to minimize the expected arrival time variation due to these variables. The authors of [5] proposed a real-time autonomous accident detection system based on computer intelligence technology. The proposed system uses k-nearest neighbor (KNN), regression trees, and feed-forward neural network models to predict the likelihood of accidents.

Finally, in [6], the authors implemented various classification models, such as logistic regression, artificial neural networks, decision trees, k-nearest neighbors, and random forests, to predict incident severity and demonstrate the performance of the proposed model with significant accuracy through experimental results. They also implemented a web-based message alert system to warn users through IoT devices.

3.1.2. Traffic Flow Prediction

Congestion prediction is one of the most common subjects in smart mobility research because, with the advent of connected vehicles (CV), the importance of research has increased as vehicles share their coordinates. This section includes an introduction to research on congestion prediction, optimal route selection, and arrival time prediction.

The authors of [7] focused on traffic delay factors. The multilayer perceptron (MLP), convolution neural network (CNN), long short-term memory recurrent neural network (LSTM RNN), gated recurrent unit recurrent neural network (GRU RNN), and autoregressive integrated moving average (ARIMA) methods were used for smart mobility data analysis. In [8], the authors predicted traffic flow, exploiting the stacked autoencoder (SAE) model. The proposed method was compared with backpropagation, random walk, support vector machine (SVM), and radial basis function (RBF) methods.

In addition, in [9], the authors proposed prediction of vehicle speed by exploiting route types, route curvature, driver behavior, weather, and traffic condition data. The model uses big data analytics and an adaptive neuro-fuzzy inference system (ANFIS). In [10], the authors proposed a system to reduce travel time and CO_2 emissions using connected vehicles to detect traffic congestion and allow users to change the paths to faster routes. A Cassandra-based big data cluster was developed, and simulation of urban mobility (SUMO) and OMNet++ discreet event network simulators were used to evaluate the proposed system. Moreover, in [11], the authors proposed a fuzzy theory and an SVM-based real-time highway traffic congestion prediction model. The model exploits Apache Storm's spout and bolt using road density, traffic condition, and rainfall data. The authors of [12] investigated the feasibility of an active real-time traffic monitoring strategy that simultaneously evaluates operation and safety. A real-time collision prediction model was implemented using data mining algorithms, such as random forest and Bayesian inference techniques.

Furthermore, the authors of [13] proposed strong reliance on results from big data prediction systems that predict accidents, estimated arrival time, and custom clearance times. Accident analysis and research revealed that driver mistakes are the main cause of car accidents. Fully autonomous vehicles eliminate human error and reduce accidents. The authors proposed a very light and simple algorithm for identification and calculation of optimal trajectories. A framework for calculation of safety indices and travel time was proposed in [14], exploiting real-time data collected from connected vehicles. The results revealed that the proposed algorithm can provide different paths when considering the safety index and travel time.

Moreover, in [15], the authors implemented an artificial neural network (ANN) composed of three layers to predict traffic flow. The ANN was trained to exploit forward propagation and backpropagation learning according to the date, time, traveler's vehicle, commercial vehicles, number of vehicles occupying the road segment, and vehicle speed. A CNN, Region-CNN (R-CNN), Fast R-CNN, and Faster R-CNN were used in [16] to recognize vehicle models or identify vehicle classifications and locations. These methods focus on building a current road traffic situation recognition system. A graph was generated using traffic data observations (TDOs) to determine the historical average of the traffic flow in [17]. Finally, regression models, such as linear regression, sequential minimal optimization regression, and M5-based regression trees, were used to determine traffic data and annual average daily traffic (AADT).

A multilevel strategy was introduced in [18] to process raw big data into a compressed time series. This strategy employed Granger causality and Lasso regression, which are suitable for causal and regression analysis, respectively, to better prepare the data for traffic modeling. In addition, an objective methodology based on engineering judgment considering the traffic accident rate of the road section, the degree of roadside development, and the geometric characteristics of the road section on urban roads was developed in [19].

Furthermore, the authors of [20] discussed an extended smart traffic management platform (STMP) based on unsupervised online incremental machine learning, deep learning, and deep reinforcement learning. STMP integrates heterogeneous big data streams, such as IoT, smart sensors, and social media, to detect conceptual drift; distinguish between repetitive and non-recurring traffic events; and determine impact propagation, traffic flow prediction, commuter sentiment analysis, and optimized traffic control. In [21], the authors aimed to develop tools for accurate and timely prediction of traffic flow information. The authors proposed a traffic flow prediction system with reduced complexity using machine learning, genetic algorithms, soft computing, and deep learning algorithms.

In [22], the authors presented a decision support system (DSS) framework that proposes alternative travel strategies for car use by assessing large volumes of traffic system data from different devices. The proposed framework was structured to collect, integrate, aggregate, converge, manage, and disseminate open big data. The framework architecture is based on a centralized database management system (DBMS) aimed at performing geographic data mining analysis. In addition, a regression model based on LSTM to predict 4 h of traffic data was proposed in [23]. The proposed system collects 24 h of traffic data online and labels it to construct an LSTM model. The proposal proves its efficiency by comparing the performance to logistic regression. The authors of [24] exploited a deep neural network (DNN) to classify traffic congestion in non-congested and congested areas. The ANN was composed of three layers, and the results indicated 99% accuracy.

Nei-TTE, a neighborhood-based deep learning method for travel time estimation, was discussed in [25]. Nei-TTE divides the entire trajectory into several separate segments and uses historical trajectory data that are perfect at the time level. The proposed model captures the characteristics of each segment and uses the trajectory characteristics of adjacent segments when the road network topology and speed interact. In [26], the authors proposed intelligent transportation based on real-time traffic conditions using graphs. They suggested an architecture that efficiently processes real-time vehicle big data using a parallel processing system and big graph processing technology. In addition, various graph algorithms were used to intelligently respond to user queries.

A traffic image analysis system based on computer vision techniques was proposed in [27]. The system core detects and classifies vehicles in traffic images. Thus, two models were implemented. The first model is an MoG and SVM system, and the second is based on Fast R-CNN for object detection in images.

Furthermore, a general architecture for fog computing applications and services in a vehicle ad hoc networks (VANET) environment was proposed in [28]. A proof-of-concept system was provided for data analysis in a hybrid VANET/fog environment. The system was used in two fog applications: one for detection of road traffic anomalies and the other for prediction of bus arrival times. A graph-oriented mechanism for implementation of smart city transportation systems was proposed in [29]. Deploying sensors on roads for traffic data collection generates numerous IoT big data, and an efficient architecture was proposed using Giraph tools and parallel processing servers to generate big graphs.

In [30], the authors proposed an optimized prediction algorithm for a radial basis function neural network based on an improved artificial bee colony (ABC) algorithm in a big data environment. Due to the uncertainty of traffic flow, the existing linear model had difficulty eliciting user satisfaction. The proposed method adopts the ABC algorithm and analyzes a nonlinear time series to optimize the weights and thresholds of the radial basis function (RBF) neural networks. Therefore, this model performed better than the KNN, RBF neural network, IBP neural network, and CARBF neural network models. The author's goal in [31] was to propose a new statistical modeling method that determines the optimal historical dataset through various analyses for each link and provides more accurate traffic flow predictions by day of the week. A three-step filtering algorithm based on change point analysis, correlation analysis, and Monte Carlo simulations was suggested to determine the optimal historical date range.

Finally, Sipresk, a platform that analyzes urban traffic data to gain insight into traffic patterns, was suggested in [32]. Sipresk consists of data, analysis, and management layers and has several verified use cases, such as finding the average speed and congestion sections on specific highways.

3.1.3. Predict and Minimize Transportation Emissions

Many cars still use fossil fuels. Fossil fuels are essential to our lives, but they pollute the environment. In this section, we review papers investigating prediction and minimization of vehicle emissions.

In [33], the authors proposed a three-layer perceptron neural network to learn and predict carbon emissions. The authors exploited five features: global positioning systems (GPS), carbon emissions, roads, POIs, and meteorology. The prediction results were good compared to Gaussian naive Bayes, linear regression, logistic regression, stacked denoising autoencoder, and deep belief network. Moreover, the authors of [34] estimated fuel consumption and emissions by exploiting GPS data. They proposed an N-dimensional representation to analyze and visualize GPS data. Another study [35] demonstrated that pavement quality, traffic volume, and the climate of roads affect energy consumption and CO_2 emissions. The proposed ranking algorithm was called recommend paths.

Furthermore, the authors of [36] developed an urban atmospheric dataset by installing a pollution measurement platform on Google Street View vehicles and repeatedly sampling all distances in an area of about 30 km^2. An analysis of these collected data revealed that the map of annual weekly NO, NO_2, and black carbon results had a steeply persistent pattern of contamination of up to 5 to 8 times. The authors of [37] proposed a system that recommends routes and vehicle speeds that minimize NOx emissions and use on-demand route optimization. Finally, In [38], a spatial econometric model was used to study the impact of smart transportation on CO_2 emissions in China.

3.2. Big Data Framework

Many big data frameworks are available to store and analyze the growing amount of big data. In this section, we discuss approaches to architectures that effectively store and analyze smart mobility big data using various big data frameworks, such as Apache Hadoop and Apache Spark.

- MapReduce is a software framework released in 2004 by Google for distributed parallel computing and large-capacity processing consisting of a mapper and a reducer. The mapper uses a key-value pair to make sorting and grouping advantageous. MapReduce undergoes the process of inputting, splitting, mapping, shuffling, reducing, and outputting. When data are entered during the input process, they are split (splitting process) and stored in the Hadoop Distributed File System (HDFS), and the map function results are combined during the shuffling process, which is the intermediate stage of the mapping and reducing tasks. Finally, in the reducing step, all values are integrated to output the results.
- Apache Hadoop is an open-source big data processing framework. It was developed to process and store increasing amounts of data, given the growth of search engines such as Yahoo and Google, and to provide users with a fast response speed. Starting with the HDFS and MapReduce framework, it expands and develops in the sense that it includes the entire Hadoop ecosystem, including data storage, execution engine programming, and data processing.
- Apache Spark is a unified big data analysis framework with various built-in modules, such as SQL, streaming, machine learning, and graphing. It can run on Apache Hadoop, Apache Mesos, and Kubernetes. The SQL component can show tables such as relational databases (RDBs) and queries in SQL style. The streaming component can apply real-time data to applications. The machine learning component, MLlib, includes various machine learning models that users can use easily to edit hyperparameters. Finally, the graph component, GraphX, is Apache Spark's API for graph and graph-parallel computation. It is flexible and works seamlessly with both graphs and collections. In addition, GraphX comes with a variety of graph algorithms. It competes with the fastest graph systems in terms of performance while retaining Spark's flexibility, fault tolerance, and ease of use.
- Apache Kafka is an event streaming platform widely used to collect, process, and store streaming events or general data without a separate start or end. In addition, it is an open-source distributed event streaming platform for high-performance data pipelines, data integration, and mission-critical applications.

The authors of [39] developed Marmot, an extension of the existing Apache Hadoop, to facilitate the implementation of geospatial big data applications in the MapReduce framework. Marmot is an architecture that processes geospatial big data in parallel using MapReduce. Marmot's data structure closely matches the relational database management system (RDBMS), where a record in Marmot corresponds to a record in the database system, with the smallest unit of data elements comprising one or more column values in a table. Marmot has two types of RecordSetFunctions: spatial and non-spatial operators. Spatial operators input, analyze, and output spatial data. Marmot exhibited superior execution speed compared to other models, including the existing spatial Hadoop.

As reported in [40], text recognition architectures existed in the past but were limited by factors such as slow speed and low accuracy due to high computational volume. Therefore, a study was conducted to improve text recognition accuracy with a low computational volume using MapReduce and Apache Hadoop to recognize the unique code written on the container. The proposed method first recognizes the unique code of the container with a camera and stores it in HDFS. Second, it converts it to a 20 $\times$ 20 gray image. Third, the text is recognized by reducing each node's computation volume using MapReduce optical character recognition (OCR). Then, it recognizes each character instead of recognizing the characters as a sentence or word and merges the last recognized characters. Through this process, it is possible to analyze big data efficiently.

The authors of another study [41] proposed a comprehensive and flexible architecture based on a distributed computing platform for real-time traffic control. Apache Kafka was used for the data pipeline and stream processing, and data were stored in HDFS. In addition, SUMO was used to simulate the opening and closing control of the highway shoulder to prove the feasibility of the proposed method.

The authors of [42] proposed an efficient model using resources from autonomous vehicles. This architecture consists of a distributed data storage mechanism for real-time analysis and an in-vehicle cloud server tool for batch processing of offline data. In addition, a workflow model for big data architecture was designed to inspect streaming data in real time.

Furthermore, the authors of [43] exploited a vehicle network without installing any additional infrastructure or hardware. An optimally distributed data-hopping mechanism was proposed that enables delay-tolerant data routing through a network of connected vehicles. The authors formulated a Markov decision process to solve the next-hop decision optimization and computational complexity problems, proposing a heuristic algorithm and demonstrating performance improvement through extensive simulations.

In [44], an intelligent smart transportation system with big graph processing using the Hadoop ecosystem was proposed using Apache Spark rather than Apache Hadoop to increase system efficiency in big data processing. Metro graph processing was performed using Giraph through Apache Hadoop. The proposed architecture exhibited efficiency in terms of scalability and real-time data processing.

The authors of another study [45] proposed a decision support system (DSS). In this architecture, public transport is promoted for those who use their cars by proposing a DSS to help public transport planners make decisions related to sustainable mobility development.

The authors of [46] proposed an architecture that calculates the quality of a public transportation system using the MapReduce method after preprocessing the GPS data. By applying the MapReduce paradigm implemented in the Apache Hadoop framework, input data not useful for statistical calculation are filtered out, and useful records are sorted and classified. In addition, based on the ideas presented in related literature, the authors proposed a method for accurately predicting travel destinations based on smart card data. This estimate allows one to calculate demand and OD matrices important for transportation planning that are difficult to obtain using traditional methods.

The authors of another study [47] presented a method to build a smart transportation system using big data analytics. Billions of connected devices generate terabytes of vehicle network data. These big data can make smart cities a reality by analyzing many aspects of transportation. This architecture has four layers. The first layer is the input layer, including the collection of various types of data, such as from security surveillance cameras, smart traffic signals, GPS, radio-frequency identification (RFID), and the Internet. The second layer is the storage layer, where data collected from various sources are stored for analysis. The third layer is the analysis layer, which takes data from the storage layer and analyzes them using various data analysis tools in a parallel or grid computing environment. Finally, in the communication layer, analysis results are provided to users in real time using other communication media, such as mobile, radio, or television networks or the Internet.

The authors of [48] suggested processing the big data architecture of autonomous and connected vehicles driving sensor data. The proposed architecture is an improved model of the existing Cooperation Intelligence Transport System (C-ITS). The architecture is based on the Hadoop ecosystem.

Furthermore, the authors of [49] focused on the big data architecture of ITS. This architecture exploits Apache Hadoop and Apache Spark. It stores traffic management and traffic analysis data for dynamic toll charging. The authors of [50] discussed an architecture capable of built-in storage and analysis. The proposed architecture comprises four modules: a big data collection and preprocessing unit, a big data processing unit, a big data analysis unit, and a visualization unit. This architecture was designed to collect, store, analyze, and visually present continuously produced ITS data to the users.

In [51], the authors proposed, developed, and applied a microservice-oriented big data architecture (MOBDA) to meet the functional and non-functional service performance requirements of ITS. This architecture was proposed to combine stream processing and batch processing of big data for smart computation of microservice-oriented transportation metrics that can serve the various needs of users.

The authors of [52] proposed a framework for acquiring data and predicting the latency of buses using Bayesian networks after preprocessing. In addition, the authors of [53] suggested a fog-computing-based ITS data analysis architecture. The proposed method exploits both fog and cloud computing. Thus, this architecture compensates for its drawback by providing faster data processing than existing cloud computing frameworks. A large-scale data analysis architecture was suggested using a wireless sensor network, big data, data mining, and other methods in [54]. The proposed architecture can ensure the compatibility of terminals, the usability of functions, real-time data, the universality of regions, and the accuracy of information, providing personalized traffic information service for users.

The authors of [55] used MapReduce to preprocess the driving data method. The authors investigated five preprocessing methods: range filtering, meaningless value exclusion, variable filter comparison, statistical technique application, and driving pattern search.

Big data technologies such as Spark on Hadoop and MongoDB can process real-time and historical data [56]. This study also highlights the influence of Spark cluster execution. The efficiency of Spark clusters in standalone mode can be observed by comparing the test results. A distributed cloud computing framework based on a big data approach was proposed to scale storage and computing resources to collect and process traffic from large networks in a reasonable amount of time [57]. In addition, a new framework that focuses on real-time anomaly detection based on big data technology was proposed in [58]. This framework consists of BroIDS, Flume, Kafka, Spark Streaming, SparkMLlib, Matplot, and HBase, which were evaluated to demonstrate their efficacy, especially in terms of accuracy, memory consumption, and execution time.

Table 1 summarizes the architectures used for big data frameworks. Many subjects are considered, such as storage of geospatial data, traffic control, optimization of vehicle pacing, path optimization, and security. Nevertheless, none of the listed architectures deals with user safety or environmental problems.

Table 1. Summary of the literature review on big data architectures.

S No.	Paper (Year)	Focus
1.	Jo et al. (2018) [39]	• Stream-based system that extends and supports Apache Hadoop; • Seamless integration between spatial and non-spatial operations; • Supports the automatic construction of MapReduce tasks.
2.	Ayed et al. (2015) [40]	• Recognizes unique code written on the container; • Increases text recognition accuracy; • Monitors the moving trajectory of cargo in real time; • Uses MapReduce and Apache Hadoop.
3.	Amini et al. (2017) [41]	• Based on a distributed computing platform; • Real-time traffic control.

Table 1. *Cont.*

S No.	Paper (Year)	Focus
4.	Daniel et al. (2017) [42]	• Distributed data storage mechanism; • Vehicle cloud server tool; • Analysis algorithm for batch processing of offline data; • Effectively obtains resources from autonomous vehicles.
5.	Si et al. (2016) [43]	• Optimal distributed data-hopping mechanism; • Enables delay-tolerant data routing; • Connected vehicle network.
6.	Rathore et al. (2018) [44]	• Ensembles smart ITS and big graph processing; • Uses the Hadoop ecosystem.
7.	Fiore et al. (2019) [45]	• Helps determine travel strategy; • Helps transit managers determine transit routes.
8.	Nesmachnow et al. (2017) [46]	• Estimates the quality of a public transportation system; • Uses MapReduce.
9.	Shukla et al. (2016) [47]	• Presents a method to build smart transportation systems; • Analyzes various data; • Security surveillance cameras, smart traffic signals, GPS, RFID, and the Internet.
10.	Aelee et al. (2020) [48]	• Improves the existing C-ITS by using the Hadoop ecosystem; • Designed for autonomous vehicles and connected vehicles.
11.	Mounica et al. (2019) [49]	• Analyzes data from various data sources above the ITS; • Collects data and uses dynamic toll-charging systems.
12.	Gohar et al. (2018) [50]	• Built-in storage; • Collects, preprocesses, stores, analyzes, and visualizes data; • Uses Apache Hadoop.
13.	Suriya et al. (2020) [51]	• Service-driven modeling for ITSs; • Adopts microservice- and component-based hybrid technology.
14.	Jackson et al. (2021) [52]	• Bayesian framework; • Predicts bus latency.
15.	Darwish et al. (2018) [53]	• Fog-based big data analytics architecture; • Efficient big ITS data processing.
16.	Liu et al. (2018) [54]	• Data analysis architecture; • Wireless sensor network, big data, data mining, and other advanced technologies; • Personalized traffic information service.
17.	Cho et al. (2017) [55]	• Proposes a preprocessing method using the MapReduce mechanism; • Filters ranges, excludes meaningless values, compares filters from variables, applies statistical techniques, and finds driving patterns.
18.	Guerreiro et al. (2016) [56]	• Efficient Spark cluster; • Combines Spark on Hadoop and MongoDB.
19.	Laboshin et al. (2017) [57]	• Large-scale traffic network data; • A distributed cloud computing framework.
20.	Ariyaluran et al. (2022) [58]	• Real-time anomaly detection; • Combines BroIDS, Flume, Apache Kafka, Apache Spark, Matplot, and HBase.

In this section, we discuss many big data frameworks, including Apache Spark, and Kafka. With so many different types of big data frameworks, it is important to know the strengths and weaknesses of each framework and use them in harmony as you progress through your research. For example, Apache Spark streaming is slower than Apache Kafka streaming, which is processed on a per-data-stream basis because real-time data are broken into microbatches, then processed. However, it has the advantage of good compatibility with other systems.

3.3. Summary

Section 3.1 introduced studies involving prediction of traffic flow, traffic accidents, and emissions using artificial neural networks and machine learning algorithms such as KNN, linear regression, and random forest. RNNs and LSTM are suitable for prediction of time series data. Fast R-CNN and Faster R-CNN, which improve the computational bottleneck of existing CNNs, are also commonly used.

Some of the key gaps identified in the reviewed literature include:

- Data standardization issues;
- Privacy of users;
- Sparse data and dense data issues;
- How to use emission prediction results.

In the reviewed papers, Apache Kafka and Apache Spark streaming were primarily used for real-time data analysis [41,42,54,58]. Apache Spark streaming has slower query processing times compared to Apache Kafka, but it is more compatible, so it is important for researchers to decide which model is most appropriate for their research. The incremental problem of big data has always been a concern. Multiple authors [39,40,47,55] have improved the incremental problem of big data by using a big data framework.

4. Application of Smart Mobility with Big Data

4.1. Security

Data generated by smart mobility sometimes contain personal or sensitive information, so security is emphasized. There is also a risk of exposure while storing the generated data on a data server in real time. In this section, we review possible security issues that may arise between data transmission, referencing papers on the security of big data systems.

For example, the authors of [59] suggested a secure data collection method when generating data using the Internet of Vehicles (IoV). Collected data are stored in HDFS with three components: the vehicle node, sink node, and big data center. The node sends the data to the sink node to reduce storage. Encryption and decryption are performed using the symbol *key_vc*, which is generated by a single sign-in algorithm. Next, the sink node checks the *key_vc* and allows access to the big data center if appropriate. The authors of another study [60] proposed an automated, secure continuous cloud service availability framework for smart connected vehicles, enabling intrusion detection mechanisms against security attacks and providing services that meet users' quality-of-service (QoS) and quality-of-experience (QoE) requirements. The proposed framework detects attacks through data traffic analysis, reduction, and classification techniques for requests that may occur during intrusions.

The authors of [61] focused on the privacy of ride sharing. Ride-sharing companies represent a risk of personal information infringement because they require users to disclose sensitive details about travel time and routes, as well as their pickup and dropoff locations. The authors proposed an encryption method using binary vectors. The authors of another study [62] found that collecting real-time traffic situation data using a connected vehicle and offloading them wirelessly to an edge computing device (ECD) increases the risk of personal information leakage, resulting in tracking, identity modulation, and virtual vehicle hijacking. The authors applied the non-dominated sorting algorithm-2 (NSGA-2) to realize multipurpose optimization to reduce execution time and energy consumption and prevent privacy conflicts.

The authors of [63] proposed security- and privacy-based access control (SPBAC) for connected vehicles. The proposed model allows security officials to access information, privileges, and roles rather than only roles for officials traveling in vehicles with the same fleet. Furthermore, the authors of [64] proposed a mechanism for randomization of data streams to ensure personal privacy. The proposed mechanism is data-adaptive and adjusts the noise for data correlations. The authors of another study [65] suggested a threshold-based image extraction solution for ITS and proposed an algorithm that uses Fast R-CNN to divide images according to importance levels and protect personal information using high thresholds for sensitive parts.

The authors of [66] suggested a blockchain framework to address privacy and security issues associated with big data and smart mobility. The proposed plan is called blockchain smart mobility data (BSMD) and is built according to the following principles: (a) Regarding user privacy, multiple DIDs are assigned to a single user, making it difficult to anonymize information and associate data. (b) Data ownership allows users to own a digital identifier and cancel the connection. (c) Concerning data transparency and audibility, anyone can access the ledger, and nodes can track all transactions related to the information. (d) For granular access control, smart contracts define what information nodes share. The authors of another study [67] proposed a multilayer blockchain framework for BSMD to address privacy, security, management, and scalability issues. It was developed as a six-layer model that defines the flow of mobility information, and open-source code is available in the transportation community. Finally, the authors of [68] focused on personal information issues that stand out as unmanned aerial vehicles (UAVs) are developed. The authors applied a number theory research unit cryptosystem to encrypt blockchain data.

The future of mobility will be connected to the Internet, and security threats such as hacking and privacy breaches will be a prominent concern. According to global market research firm MarketsAndMarket, the global automotive cybersecurity market is expected grow from USD 2 billion in 2021 to USD 5.3 billion in 2026 [69]. As a result, not only automakers such as Hyundai Motor Company but also IT companies such as LG Electronics have recognized the importance of automotive cybersecurity and are developing software and technologies to enhance security.

4.2. Visualization and Monitoring

Visualization of data analysis results is important because it is intuitive and clear to non-experts. The importance of data visualization is emphasized as the amount of data increases and becomes more complex. Visually checking data analysis results has advantages in terms of understanding and time efficiency. According to [70], the speed of information processing varies greatly from one sense to another and compared to a computer system.

For example, the authors of [71] developed a visualization and analysis method for transportation data, exploiting taxi and bus GPS and bicycle-sharing data. The proposed application is a python package, TransBigData, which was used by data science researchers and engineers, and through spatiotemporal transportation data analysis, government and other enterprises can support efficient and reliable management decisions. Furthermore, the possibility of motivating and improving traffic operation highway safety by exploiting real-time microwave vehicle detection system (MVDS) data was studied in [12].

The authors of another study [17] presented a Java-based traffic information observation (TDO) tool to filter and visualize historical traffic big data. The authors of [72] established a solution to create a platform for pothole detection and road monitoring, implementing road observers using IMU sensors attached to embedded systems and proposing smart environmental monitoring and analysis, an IoT-based monitoring platform system, and an application that transmits real-time big data. The authors of [73] proposed a system that uses information visualization techniques to analyze urban traffic data and the impact of emissions on urban air quality. The study showed that citizens and public agencies can use effective visualization to identify trends, detect congested road sections at certain times, and visualize the correlation between road traffic and air pollution for system users. The authors of [74] used an ANN with fuzzy logic to determine the remaining construction life (LCL) of a permanent magnet (PM) by mapping the source swell and the time at which the electric vehicle is driven and outputting the external temperature, twisting temperature, the moment of component lubrication, and the split weakening of the responsible magnetic field in the context of connected vehicles. The authors also proposed a monitoring system whereby cloud computing, traffic data, and intelligent transportation systems (ITS) can improve PHEV energy management.

Finally, the authors of [75] analyzed road traffic and pollution data and applied algorithms that leverage the MapReduce framework running on Hadoop clusters. Results were collected, and the path with the least pollution was calculated and visualized.

4.3. Connected Vehicles

Connected vehicles can connect to nearby devices using wireless internet. Connected vehicles are an important factor in the IoT industry.

As many sensors are attached to connected vehicles, more data are generated, and not only are location and speed shared, but data collection and analyses are also conducted. The collected data are exploited for traffic flow detection, faster path recommendation, and reduced emissions. Figure 2 presents the concept of a connected vehicle. The connected vehicle is associated with many objects, generating considerable data. In this section, we review the applications of data generated by connected vehicles in depth, as well as research conducted to date.

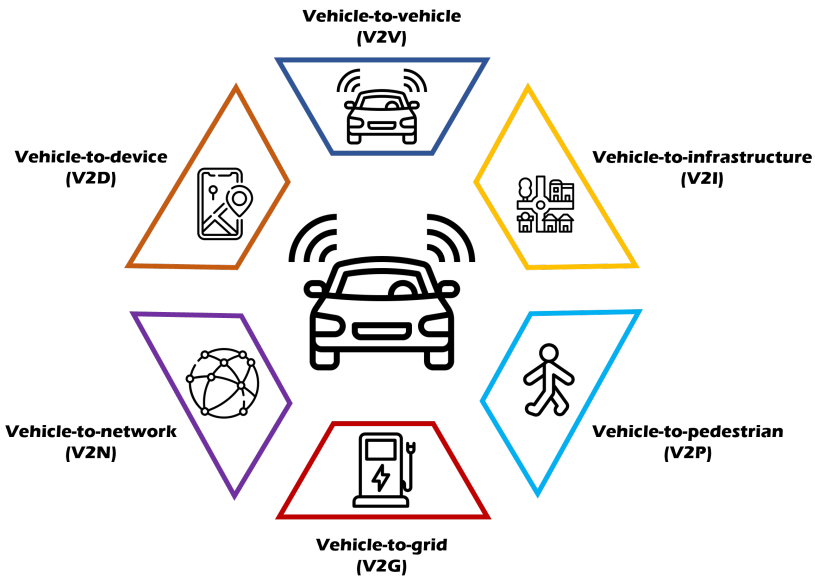

Figure 2. Connected vehicle concept.

The authors of [15] introduced the concept of data management for vehicles by proposing an onboard data management layer so that the vehicle can serve as a data platform that can store, process, and share data. Moreover, the authors of [76] proposed ThinGs in a fog (TGIF), a system designed to support interdisciplinary research within a wide range of IoT applications. The authors of [77] reported on the open-source Pikalert® system, which combines weather information with real-time data from connected vehicles to provide important information to improve the safety and efficiency of ground transport systems. This framework can be applied to a wide variety of user communities and is designed to quickly collect datasets as they become available. In addition, the authors of [78] proposed an architecture of connected vehicle platforms that can efficiently structure and analyze data by collecting raw data generated by various devices. They also proposed an adaptive memory-based data storage algorithm to be applied to the data reception layer of connected vehicle platforms.

Furthermore, the authors of [79] proposed a distributed intelligent transportation system that applies intelligence to connected vehicles in dynamic decision making to pass through specific areas such as rotations and intersections. This transportation system assumes that each connected vehicle is an ant and focuses on dynamic decision making, exploiting the ACO (Ant Colony Optimization) algorithm. Moreover, the authors of [80] proposed an adaptive signal control optimization method aimed at minimizing the time to board vehicles by taking advantage of the fact that connected vehicles can easily use each other's vehicle location, speed, and other traffic information.

The authors of [81] developed a CAN (controller area network) bus and a 4G–LTE-based module that transmits vehicle powertrain information by utilizing the connectivity and network of connected cars. The proposed system can check vehicle failure with a mobile application by exploiting the manufacturer's diagnostic code. A new modular FL (federated learning) method was studied in [82] to enhance road user/object classification based on LiDAR data for the integration of connected car decision making.

A methodology for quantification of changes in vehicle motion and for the study of instantaneous driving behavior for collision accident types at intersections was presented in [83], using longitudinal and lateral variability. More than 125 million basic safety message data transmitted between more than 2800 connected vehicles were analyzed and integrated with historical crash and road inventory data at 167 intersections. Furthermore, the authors of [84] proposed an easy-to-implement origin and destination (OD)-based segmentation technique to improve trunk signal adjustment by utilizing automatically collected vehicle trajectory data from connected vehicles.

4.4. Autonomous Vehicles

In this section, we review studies on how autonomous vehicles select routes, considering limitations such as safety and minimum time. An autonomous vehicle can operate on its own without driver or passenger manipulation. The emergence of self-driving cars is expected to reduce traffic accidents and retaliatory driving due to driver carelessness. However, there are concerns that this usage could reduce jobs for people in the transportation industry and increase cyber attacks.

In October 2016, the National Highway Traffic Safety Administration (NHTSA) officially adopted the autonomous driving stage released by the SAE, which is currently divided into six stages, as listed in Table 2.

Table 2. Six levels of autonomous driving technology as defined by the SAE.

Stage	Feature	Contents
Level 0	No automation	The driver controls all operations and encourages all dynamic driving.
Level 1	Driver assistance	The vehicle is controlled by the steering assistance system or the acceleration/deceleration assistance system, but the person performs all functions for dynamic driving.
Level 2	Partial automation	The vehicle is run by a steering support system or acceleration/deceleration support system, but the driving environment is monitored by the person, and the driver is responsible for safe driving.
Level 3	Conditional automation	The system controls all aspects of operation, but if the system requests driver intervention, the driver must properly control the vehicle, and the driver is responsible for safe control.
Level 4	High automation	Key driving controls, driving environment monitoring, and emergency responses are all conducted by the system but not always entirely controlled by the system.
Level 5	Full automation	The system is responsible for driving in all road conditions and environments.

For example, one study [85] predicted that self-driving cars would evolve into services, not products, through artificial intelligence and big data analysis. An online autonomous vehicle management system using network calculus (NC) was proposed to provide the best service to users.

Another study [86] proposed an Apache Kafka and Spark-based autonomous vehicle development solution that can collect various data in real time. Therefore, it is expected that future cars will be able to select the optimal route quickly and autonomously. The authors of [87] proposed a model that solves internal algorithm performance and data storage and processing problems when collecting, analyzing, and storing data generated by autonomous vehicles. The model consists of Flume, Apache Kafka, and Zookeeper for collection and transmission of data to the storage system.

The authors of [88] suggested an optimization method to improve the traffic efficiency of autonomous vehicles in conflict areas. In addition, the authors of [89] emphasized the importance of autonomous vehicles in determining mandatory lane changes (MLC) in urban road networks because incorrect MLC decisions on main roads threaten efficiency and stability. Therefore, the aim of the study was to solve two key strategic decision variables through simulation experiments.

The authors of [90] proposed a sensor fusion mechanism that combines resources acquired from autonomous vehicles such as 3D camera sensor data and LiDAR sensor information to provide an optimized solution for route selection. Moreover, the authors of [91] discussed an automatic parking system exploiting sensors of autonomous vehicles such as radar, cameras, and ultrasonic sensors. In addition, the authors of [92] proposed an IoT gateway and deep learning-based self-driving car ISS (integrated self-diagnosis system), that collects information from sensors of autonomous vehicles and analyzes self-diagnosis and their effects. The proposed system uses deep learning to deliver information after self-diagnosis and manage the time when parts of autonomous vehicles can be safely replaced.

Furthermore, in [93], a system was proposed to improve the control design of autonomous vehicles through big data analysis of signals measured in autonomous vehicles. The efficiency of the proposed control system was demonstrated through complex simulation examples using the CarSim simulation environment.

In addition, the authors of [42] proposed an architecture that stores distributed data, processes streaming data for real-time analysis, and processes batch offline data to effectively utilize data collected by autonomous vehicles. The authors also described the data classification algorithm of distributed storage devices and mathematical modeling to analyze the data classification function of autonomous vehicles.

The authors of [94] proposed an autonomous vehicle dynamics analysis focusing on side stability. The analysis was based on the collection of big data from vehicle signals. The key to the proposed analysis method is the C4.5 machine learning algorithm. The purpose of inspection is to analyze the relationship between the various signals (e.g., yaw rate, side slip angle, longitudinal speed, and adhesion coefficient) and the lateral dynamics of the vehicle. In addition, the authors of [95] proposed an OLAP-based (online analytical processing) analysis tool that utilizes big data to improve the reliability and efficiency of decision making for autonomous vehicles.

The authors of [96] proposed an autonomous driving framework using deep reinforcement learning. This framework integrates the attention model to improve the time complexity by using glimpses and action networks to direct the CNN kernel to the input area associated with the driving process. Moreover, the authors of [97] proposed an application of the deep reinforcement learning technique to determine optimal driving policies by maximizing long-term rewards in interactive environments to solve complex control problems. The proposed architecture uses LSTM to model an interactive environment delivered to a DQN (Deep Q-Network), including historical driving information.

In [98], a new probability recognition algorithm is presented as a real-time solution for data connection, object tracking, and object classification for autonomous ground vehicles under all weather conditions. The proposed method implements a state-of-the-art vision detection algorithm that includes directional information for autonomous vehicle applications. Moreover, in [99], unmanned autonomous vehicles were proposed with a focus on enhancing traffic sign recognition performance to assist in autonomous driving. Because not only color but also time constraints are important when recognizing traffic signs on a video basis, the authors proposed an algorithm to recognize traffic signs by fusing space and time functions with CNN bases. The authors of [100] proposed four rules for autonomous vehicles to prioritize ethical data exploitation and decision making between pedestrians and passengers.

5. Open Challenges

In this section, we discuss the limitations of the literature review and suggest directions for future research.

Technical limitations: The limitations of studies using machine learning are that they may produce incorrect results in untrained situations, that is, they may be vulnerable to unexpected situations during the operation of the vehicle. Additionally, most machine learning models are black boxes, which makes them difficult to interpret. In order to improve the accuracy of the model, it is sometimes necessary to modify the model, but most machine learning algorithms are generated as a result of learning, and it is difficult for humans to understand the process whereby results are derived, which limits the improvement of the model. Therefore, it is valuable to develop a model that allows people to easily understand why the AI model made a certain decision and how it works, such as XAI (explainable AI).

In addition, in the case of big data, the amount of data accumulates at a rapid pace in hyperconnected systems. As a result, query response times are sometimes delayed by unnecessary exploration when making decisions. Therefore, it is very important to study how to effectively determine and explore only the necessary data.

Legal limitations: As the concept of smart mobility has not been around for a long time, the system has not yet been firmly established. The active cooperation and participation of governments, companies, users, and related communities are required for commercialization. Therefore, it seems necessary to establish technical standards and legislation that specify punishments and responsibility for accidents or illegal acts caused by errors in the artificial intelligence embedded in smart mobility devices. For example, if an accident occurs due to a central line violation in a self-driving car, it is necessary to establish standards for determining whether the responsible party is the manufacturer, the driver, or another party and how to divide the fault.

Ethical limitations: There are ethical dilemmas in both smart mobility and artificial intelligence research. The dilemma of "Which one should artificial intelligence choose between probabilistic choice and ethical choice?" is an age-old issue. For example, in a situation in which an elderly person and a young man are standing on either side of the road and there is no way to avoid or stop them, opinions on what choice a self-driving car with artificial intelligence should make vary greatly among people. The young man is more likely to survive, even if an accident occurs. However, in our society, protecting the more productive young man is also an option. Lastly, there is a way for the driver to be injured with neither the young man nor the elderly person being injured by colliding with a wall or guardrail. There is no correct answer to these questions. However, artificial intelligence developers feel burdened when prioritizing certain choices over ethical choices and survival rates.

6. Conclusions

Various means of transportation are connected to the Internet, and new attempts are increasing accordingly. In particular, the combination of smart mobility and advances in IoT, AI, and big data analysis technologies will change much of our mobility. In this paper, we extensively analyzed and reviewed various recent studies exploiting data generated by smart mobility. However, due to technological development, there are positive aspects such as the convenience of movement, safety improvement, fast path recommendation, and accurate arrival time prediction, as well as negative aspects, such as job loss for transportation businesses and ambiguous selection of responsible drivers and autonomous vehicle manufacturers. Therefore, it seems necessary to appropriately use technology that develops at an exponential rate that is expected to continue in the future and to change the cultural paradigm.

Many automakers are now building connected vehicles. In the near future, all modes of transportation will be connected to the Internet, sharing their location, destination, speed, and other information for various purposes, such as traffic collection and optimization. However, there is also a risk of personal information leakage that can occur when exchanging information. In addition, when autonomous driving becomes commonplace, a solution is needed to prevent car hacking, unauthorized operation, and smart key cloning. Furthermore, since connected vehicles share their location, optimization of traffic flow can eliminate the need for traffic lights. In addition to traffic flow algorithms, more accurate image processing technologies and fast communication systems are needed.

Big data and artificial intelligence are expected to make a significant contribution to the field of smart mobility. Processing large amounts of data and information collected by smart mobility devices will make the movement of users convenient and safe and optimize traffic management.

Author Contributions: Conceptualization, D.L. and J.J.J.; methodology, D.L., D.C. and J.J.J.; software, D.L. and J.J.J.; validation, D.C. and J.J.J.; formal analysis, D.L. and J.J.J.; investigation, D.L.; resources, D.L. and J.J.J.; data curation, D.C. and J.J.J.; writing—original draft preparation, D.L.; writing—review and editing, D.C. and J.J.J.; visualization, D.L. and D.C.; supervision, J.J.J.; project administration, J.J.J.; funding acquisition, J.J.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Chung-Ang University Research Grants in 2021.

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. Precedence Research. Smart Mobility Market Size to Worth around USD 250.3 Bn by 2030. Available online: https://www.globenewswire.com/news-release/2022/08/12/2497710/0/en/Smart-Mobility-Market-Size-to-Worth-Around-USD-250-3-Bn-by-2030.html (accessed on 12 January 2023).
2. Zhang, M.; Chen, C.; Wo, T.; Xie, T.; Bhuiyan, M.Z.A.; Lin, X. SafeDrive: Online driving anomaly detection from large-scale vehicle data. *IEEE Trans. Ind. Inform.* **2017**, *13*, 2087–2096. [CrossRef]
3. Park, S.h.; Kim, S.m.; Ha, Y.g. Highway traffic accident prediction using VDS big data analysis. *J. Supercomput.* **2016**, *72*, 2815–2831. [CrossRef]
4. 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; pp. 1–8.
5. Ozbayoglu, M.; Kucukayan, G.; Dogdu, E. A real-time autonomous highway accident detection model based on big data processing and computational intelligence. In Proceedings of the 2016 IEEE international conference on big data (Big Data), Washington, DC, USA, 5–8 December 2016; pp. 1807–1813.
6. Mohanta, B.K.; Jena, D.; Mohapatra, N.; Ramasubbareddy, S.; Rawal, B.S. Machine learning based accident prediction in secure iot enable transportation system. *J. Intell. Fuzzy Syst.* **2022**, *42*, 713–725. [CrossRef]
7. Chauhan, R.; Shi, Y.; Bartlett, A.; Sadek, A.W. Short-Term Traffic Delay Prediction at the Niagara Frontier Border Crossings: Comparing Deep Learning and Statistical Modeling Approaches. *J. Big Data Anal. Transp.* **2020**, *2*, 93–114. [CrossRef]
8. Lv, Y.; Duan, Y.; Kang, W.; Li, Z.; Wang, F.Y. Traffic flow prediction with big data: A deep learning approach. *IEEE Trans. Intell. Transp. Syst.* **2014**, *16*, 865–873. [CrossRef]
9. Cheng, Z.; Chow, M.Y.; Jung, D.; Jeon, J. A big data based deep learning approach for vehicle speed prediction. In Proceedings of the 2017 IEEE 26th International Symposium on Industrial Electronics (ISIE), Edinburgh, UK, 19–21 June 2017; pp. 389–394.
10. Cárdenas-Benítez, N.; Aquino-Santos, R.; Magaña-Espinoza, P.; Aguilar-Velazco, J.; Edwards-Block, A.; Medina Cass, A. Traffic congestion detection system through connected vehicles and big data. *Sensors* **2016**, *16*, 599. [CrossRef] [PubMed]
11. Tseng, F.H.; Hsueh, J.H.; Tseng, C.W.; Yang, Y.T.; Chao, H.C.; Chou, L.D. Congestion prediction with big data for real-time highway traffic. *IEEE Access* **2018**, *6*, 57311–57323. [CrossRef]

12. 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]
13. Al Najada, H.; Mahgoub, I. Autonomous vehicles safe-optimal trajectory selection based on big data analysis and predefined user preferences. In Proceedings of the 2016 IEEE 7th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), New York, NY, USA, 20–22 October 2016; pp. 1–6.
14. Hoseinzadeh, N.; Arvin, R.; Khattak, A.J.; Han, L.D. Integrating safety and mobility for pathfinding using big data generated by connected vehicles. *J. Intell. Transp. Syst.* **2020**, *24*, 404–420. [CrossRef]
15. Benaissa, K.; Bitam, S.; Mellouk, A. On-Board Data Management Layer: Connected Vehicle as Data Platform. *Electronics* **2021**, *10*, 1810. [CrossRef]
16. Zhu, Q. Research on road traffic situation awareness system based on image big data. *IEEE Intell. Syst.* **2019**, *35*, 18–26. [CrossRef]
17. Alam, I.; Ahmed, M.F.; Alam, M.; Ulisses, J.; Farid, D.M.; Shatabda, S.; Rossetti, R.J. Pattern mining from historical traffic big data. In Proceedings of the 2017 IEEE Region 10 Symposium (TENSYMP), Cochin, India, 14–16 July 2017; pp. 1–5.
18. Li, L.; Su, X.; Wang, Y.; Lin, Y.; Li, Z.; Li, Y. Robust causal dependence mining in big data network and its application to traffic flow predictions. *Transp. Res. Part C Emerg. Technol.* **2015**, *58*, 292–307. [CrossRef]
19. Kim, H.; Jung, D. Estimation of Optimal Speed Limits for Urban Roads Using Traffic Information Big Data. *Appl. Sci.* **2021**, *11*, 5710. [CrossRef]
20. Nallaperuma, D.; Nawaratne, R.; Bandaragoda, T.; Adikari, A.; Nguyen, S.; Kempitiya, T.; De Silva, D.; Alahakoon, D.; Pothuhera, D. Online incremental machine learning platform for big data-driven smart traffic management. *IEEE Trans. Intell. Transp. Syst.* **2019**, *20*, 4679–4690. [CrossRef]
21. Meena, G.; Sharma, D.; Mahrishi, M. Traffic prediction for intelligent transportation system using machine learning. In Proceedings of the 2020 3rd International Conference on Emerging Technologies in Computer Engineering: Machine Learning and Internet of Things (ICETCE), Jaipur, India, 7–8 February 2020; pp. 145–148.
22. Guido, G.; Rogano, D.; Vitale, A.; Astarita, V.; Festa, D. Big data for public transportation: A DSS framework. In Proceedings of the 2017 5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS), Naples, Italy, 26–28 June 2017; pp. 872–877.
23. Du, X.; Zhang, H.; Van Nguyen, H.; Han, Z. Stacked LSTM deep learning model for traffic prediction in vehicle-to-vehicle communication. In Proceedings of the 2017 IEEE 86th Vehicular Technology Conference (VTC-Fall), Toronto, ON, Canada, 24–27 September 2017; pp. 1–5.
24. Yi, H.; Jung, H.; Bae, S. Deep neural networks for traffic flow prediction. In Proceedings of the 2017 IEEE International Conference on Big Data and Smart Computing (BigComp), Jeju, Republic of Korea, 13–16 February 2017; pp. 328–331.
25. Qiu, J.; Du, L.; Zhang, D.; Su, S.; Tian, Z. Nei-TTE: Intelligent traffic time estimation based on fine-grained time derivation of road segments for smart city. *IEEE Trans. Ind. Inform.* **2019**, *16*, 2659–2666. [CrossRef]
26. Rathore, M.M.; Ahmad, A.; Paul, A.; Thikshaja, U.K. Exploiting real-time big data to empower smart transportation using big graphs. In Proceedings of the 2016 IEEE Region 10 Symposium (TENSYMP), Bali, Indonesia, 9–11 May 2016; pp. 135–139.
27. Arinaldi, A.; Pradana, J.A.; Gurusinga, A.A. Detection and classification of vehicles for traffic video analytics. *Procedia Comput. Sci.* **2018**, *144*, 259–268. [CrossRef]
28. Pereira, J.; Ricardo, L.; Luís, M.; Senna, C.; Sargento, S. Assessing the reliability of fog computing for smart mobility applications in VANETs. *Future Gener. Comput. Syst.* **2019**, *94*, 317–332. [CrossRef]
29. 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 International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), Bangkok, Thailand, 23–27 November 2015; pp. 512–519.
30. Chen, D. Research on traffic flow prediction in the big data environment based on the improved RBF neural network. *IEEE Trans. Ind. Inform.* **2017**, *13*, 2000–2008. [CrossRef]
31. Jeon, S.; Hong, B. Monte Carlo simulation-based traffic speed forecasting using historical big data. *Future Gener. Comput. Syst.* **2016**, *65*, 182–195. [CrossRef]
32. Khazaei, H.; Zareian, S.; Veleda, R.; Litoiu, M. Sipresk: A big data analytic platform for smart transportation. In Proceedings of the Smart City 360: First EAI International Summit, Smart City 360, Bratislava, Slovakia, 13–16 October 2015; Springer: Berlin/Heidelberg, Germany, 2016; pp. 419–430.
33. Lu, X.; Ota, K.; Dong, M.; Yu, C.; Jin, H. Predicting transportation carbon emission with urban big data. *IEEE Trans. Sustain. Comput.* **2017**, *2*, 333–344. [CrossRef]
34. Kan, Z.; Tang, L.; Kwan, M.P.; Zhang, X. Estimating vehicle fuel consumption and emissions using GPS big data. *Int. J. Environ. Res. Public Health* **2018**, *15*, 566. [CrossRef] [PubMed]
35. Louhghalam, A.; Akbarian, M.; Ulm, F.J. Carbon management of infrastructure performance: Integrated big data analytics and pavement-vehicle-interactions. *J. Clean. Prod.* **2017**, *142*, 956–964. [CrossRef]
36. Apte, J.S.; Messier, K.P.; Gani, S.; Brauer, M.; Kirchstetter, T.W.; Lunden, M.M.; Marshall, J.D.; Portier, C.J.; Vermeulen, R.C.; Hamburg, S.P. High-resolution air pollution mapping with Google street view cars: Exploiting big data. *Environ. Sci. Technol.* **2017**, *51*, 6999–7008. [CrossRef]
37. Dimokas, N.; Margaritis, D.; Gaetani, M.; Koprubasi, K.; Bekiaris, E. A Big Data application for low emission heavy duty vehicles. *Transp. Telecommun.* **2020**, *21*, 265–274. [CrossRef]

38. Zhao, C.; Wang, K.; Dong, X.; Dong, K. Is smart transportation associated with reduced carbon emissions? The case of China. *Energy Econ.* **2022**, *105*, 105715. [CrossRef]
39. Jo, J.; Lee, K.W. High-performance geospatial big data processing system based on MapReduce. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 399. [CrossRef]
40. Ayed, A.B.; Halima, M.B.; Alimi, A.M. Big data analytics for logistics and transportation. In Proceedings of the 2015 4th international conference on advanced logistics and transport (ICALT), Valenciennes, France, 20–22 May 2015; pp. 311–316.
41. Amini, S.; Gerostathopoulos, I.; Prehofer, C. Big data analytics architecture for real-time traffic control. In Proceedings of the 2017 5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS), Naples, Italy, 26–28 June 2017; pp. 710–715.
42. Daniel, A.; Subburathinam, K.; Paul, A.; Rajkumar, N.; Rho, S. Big autonomous vehicular data classifications: Towards procuring intelligence in ITS. *Veh. Commun.* **2017**, *9*, 306–312. [CrossRef]
43. Si, P.; He, Y.; Yao, H.; Yang, R.; Zhang, Y. DaVe: Offloading delay-tolerant data traffic to connected vehicle networks. *IEEE Trans. Veh. Technol.* **2016**, *65*, 3941–3953. [CrossRef]
44. Rathore, M.M.; Paul, A.; Hong, W.H.; Seo, H.; Awan, I.; Saeed, S. Exploiting IoT and big data analytics: Defining smart digital city using real-time urban data. *Sustain. Cities Soc.* **2018**, *40*, 600–610. [CrossRef]
45. 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]
46. Nesmachnow, S.; Baña, S.; Massobrio, R. A distributed platform for big data analysis in smart cities: Combining intelligent transportation systems and socioeconomic data for Montevideo, Uruguay. *EAI Endorsed Trans. Smart Cities* **2017** , *2*. [CrossRef]
47. Shukla, S.; Balachandran, K.; Sumitha, V. A framework for smart transportation using Big Data. In Proceedings of the 2016 International Conference on ICT in Business Industry & Government (ICTBIG), Indore, India, 18–19 November 2016; pp. 1–3.
48. Yoo, A.; Shin, S.; Lee, J.; Moon, C. Implementation of a sensor big data processing system for autonomous vehicles in the C-ITS environment. *Appl. Sci.* **2020**, *10*, 7858. [CrossRef]
49. Mounica, B.; Lavanya, K. Bigdata Architecture for Intelligence Transport System. *Int. J. Innov. Technol. Explor. Eng.* **2019**, *9*, 1281–1286.
50. Gohar, M.; Muzammal, M.; Rahman, A.U. SMART TSS: Defining transportation system behavior using big data analytics in smart cities. *Sustain. Cities Soc.* **2018**, *41*, 114–119. [CrossRef]
51. Asaithambi, S.P.R.; Venkatraman, R.; Venkatraman, S. MOBDA: Microservice-oriented big data architecture for smart city transport systems. *Big Data Cogn. Comput.* **2020**, *4*, 17. [CrossRef]
52. Jackson, M.D.; Leung, C.K.; Mbacke, M.D.B.; Cuzzocrea, A. A Bayesian framework for supporting predictive analytics over big transportation data. In Proceedings of the 2021 IEEE 45th Annual Computers, Software, and Applications Conference (COMPSAC), Madrid, Spain, 12–16 July 2021; pp. 332–337.
53. Darwish, T.S.; Bakar, K.A. Fog based intelligent transportation big data analytics in the internet of vehicles environment: Motivations, architecture, challenges, and critical issues. *IEEE Access* **2018**, *6*, 15679–15701. [CrossRef]
54. Liu, D. Big data analytics architecture for internet-of-vehicles based on the spark. In Proceedings of the 2018 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS), Xiamen, China, 25–26 January 2018; pp. 13–16.
55. Cho, W.; Choi, E. Big data pre-processing methods with vehicle driving data using MapReduce techniques. *J. Supercomput.* **2017**, *73*, 3179–3195. [CrossRef]
56. Guerreiro, G.; Figueiras, P.; Silva, R.; Costa, R.; Jardim-Goncalves, R. An architecture for big data processing on intelligent transportation systems. An application scenario on highway traffic flows. In Proceedings of the 2016 IEEE 8th International Conference on Intelligent Systems (IS), Sofia, Bulgaria, 4–6 September 2016; pp. 65–72.
57. Laboshin, L.; Lukashin, A.; Zaborovsky, V. The big data approach to collecting and analyzing traffic data in large scale networks. *Procedia Comput. Sci.* **2017**, *103*, 536–542. [CrossRef]
58. Ariyaluran Habeeb, R.A.; Nasaruddin, F.; Gani, A.; Amanullah, M.A.; Abaker Targio Hashem, I.; Ahmed, E.; Imran, M. Clustering-based real-time anomaly detection—A breakthrough in big data technologies. *Trans. Emerg. Telecommun. Technol.* **2022**, *33*, e3647. [CrossRef]
59. Guo, L.; Dong, M.; Ota, K.; Li, Q.; Ye, T.; Wu, J.; Li, J. A secure mechanism for big data collection in large scale internet of vehicle. *IEEE Internet Things J.* **2017**, *4*, 601–610. [CrossRef]
60. Aloqaily, M.; Otoum, S.; Al Ridhawi, I.; Jararweh, Y. An intrusion detection system for connected vehicles in smart cities. *Ad Hoc Netw.* **2019**, *90*, 101842. [CrossRef]
61. Sherif, A.B.; Rabieh, K.; Mahmoud, M.M.; Liang, X. Privacy-preserving ride sharing scheme for autonomous vehicles in big data era. *IEEE Internet Things J.* **2016**, *4*, 611–618. [CrossRef]
62. Xu, X.; Xue, Y.; Qi, L.; Yuan, Y.; Zhang, X.; Umer, T.; Wan, S. An edge computing-enabled computation offloading method with privacy preservation for internet of connected vehicles. *Future Gener. Comput. Syst.* **2019**, *96*, 89–100. [CrossRef]
63. Habib, M.A.; Ahmad, M.; Jabbar, S.; Khalid, S.; Chaudhry, J.; Saleem, K.; Rodrigues, J.J.; Khalil, M.S. Security and privacy based access control model for internet of connected vehicles. *Future Gener. Comput. Syst.* **2019**, *97*, 687–696. [CrossRef]
64. Ghane, S.; Jolfaei, A.; Kulik, L.; Ramamohanarao, K.; Puthal, D. Preserving privacy in the internet of connected vehicles. *IEEE Trans. Intell. Transp. Syst.* **2020**, *22*, 5018–5027. [CrossRef]

65. Liu, Y.; Yang, C.; Sun, Q. Thresholds based image extraction schemes in big data environment in intelligent traffic management. *IEEE Trans. Intell. Transp. Syst.* **2020**, *22*, 3952–3960. [CrossRef]
66. Lopez, D.; Farooq, B. A blockchain framework for smart mobility. In Proceedings of the 2018 IEEE International Smart Cities Conference (ISC2), Kansas City, MO, USA, 16–19 September 2018; pp. 1–7.
67. Lopez, D.; Farooq, B. A multi-layered blockchain framework for smart mobility data-markets. *Transp. Res. Part C Emerg. Technol.* **2020**, *111*, 588–615. [CrossRef]
68. Lv, Z.; Qiao, L.; Hossain, M.S.; Choi, B.J. Analysis of using blockchain to protect the privacy of drone big data. *IEEE Netw.* **2021**, *35*, 44–49. [CrossRef]
69. Global Research & Data. Automotive Cybersecurity Market by Form (In-Vehicle, External Cloud Services), Offering (Hardware & Software), Security, Application Type, Vehicle Type, Propulsion, Vehicle Autonomy, Approach, EV Application, and Region–Global Forecast to 2026. Available online: https://www.globalresearch.co.kr/report/automotive-cybersecurity-market-form-invehicle (accessed on 14 May 2023).
70. McCandless, D. *David McCandless: The Beauty of Data Visualization*; TED: New York, NY, USA , 2010.
71. Yu, Q.; Yuan, J. TransBigData: A Python package for transportation spatio-temporal big data processing, analysis and visualization. *J. Open Source Softw.* **2022**, *7*, 4021. [CrossRef]
72. Ulil, A.M.R.; Sukaridhoto, S.; Tjahjono, A.; Basuki, D.K. The vehicle as a mobile sensor network base iot and big data for pothole detection caused by flood disaster. *IOP Conf. Ser. Earth Environ. Sci.* **2019**, *239*, 012034. [CrossRef]
73. Bachechi, C.; Po, L.; Rollo, F. Big data analytics and visualization in traffic monitoring. *Big Data Res.* **2022**, *27*, 100292. [CrossRef]
74. Sreenivasu, S.; Sathesh Kumar, T.; Bin Hussain, O.; Yeruva, A.R.; Kabat, S.R.; Chaturvedi, A. Cloud Based Electric Vehicle's Temperature Monitoring System Using IOT. *Cybern. Syst.* **2023** , 1–16. [CrossRef]
75. Zenkert, J.; Dornhofer, M.; Weber, C.; Ngoukam, C.; Fathi, M. Big data analytics in smart mobility: Modeling and analysis of the Aarhus smart city dataset. In Proceedings of the 2018 IEEE Industrial Cyber-Physical Systems (ICPS), St. Petersburg, Russia, 15–18 May 2018; pp. 363–368.
76. Rayamajhi, A.; Rahman, M.; Kaur, M.; Liu, J.; Chowdhury, M.; Hu, H.; McClendon, J.; Wang, K.C.; Gosain, A.; Martin, J. Things in a fog: System illustration with connected vehicles. In Proceedings of the 2017 IEEE 85th Vehicular Technology Conference (VTC Spring), Sydney, NSW, Australia, 4–7 June 2017; pp. 1–6.
77. Siems-Anderson, A.R.; Walker, C.L.; Wiener, G.; Mahoney III, W.P.; Haupt, S.E. An adaptive big data weather system for surface transportation. *Transp. Res. Interdiscip. Perspect.* **2019**, *3*, 100071. [CrossRef]
78. Kim, C.; Choi, H.s.; Ko, J. Adaptive Memory-based Data Storage Algorithm for Connected Vehicle Platform. In Proceedings of the 2019 International Conference on Information and Communication Technology Convergence (ICTC), Jeju, Republic of Korea, 16–18 October 2019; pp. 1172–1174.
79. Bui, K.H.N.; Jung, J.J. ACO-based dynamic decision making for connected vehicles in IoT system. *IEEE Trans. Ind. Inform.* **2019**, *15*, 5648–5655. [CrossRef]
80. Yao, Z.; Jiang, Y.; Zhao, B.; Luo, X.; Peng, B. A dynamic optimization method for adaptive signal control in a connected vehicle environment. *J. Intell. Transp. Syst.* **2020**, *24*, 184–200.
81. Kwon, D.; Park, S.; Ryu, J.T. A study on big data thinking of the internet of things-based smart-connected car in conjunction with controller area network bus and 4G-long term evolution. *Symmetry* **2017**, *9*, 152. [CrossRef]
82. Barbieri, L.; Savazzi, S.; Brambilla, M.; Nicoli, M. Decentralized federated learning for extended sensing in 6G connected vehicles. *Veh. Commun.* **2022**, *33*, 100396. [CrossRef]
83. Arvin, R.; Kamrani, M.; Khattak, A.J. How instantaneous driving behavior contributes to crashes at intersections: Extracting useful information from connected vehicle message data. *Accid. Anal. Prev.* **2019**, *127*, 118–133. [CrossRef]
84. Xu, J.; Tian, Z. OD-Based Partition Technique to Improve Arterial Signal Coordination Using Connected Vehicle Data. *Transp. Res. Rec.* **2023**, *2677*, 252–265. [CrossRef]
85. Cui, Q.; Wang, Y.; Chen, K.C.; Ni, W.; Lin, I.C.; Tao, X.; Zhang, P. Big data analytics and network calculus enabling intelligent management of autonomous vehicles in a smart city. *IEEE Internet Things J.* **2018**, *6*, 2021–2034. [CrossRef]
86. Reddig, K.; Dikunow, B.; Krzykowska, K. Proposal of big data route selection methods for autonomous vehicles. *Internet Technol. Lett.* **2018**, *1*, e36. [CrossRef]
87. Kumar, S.; Goel, E. Changing the world of autonomous vehicles using cloud and big data. In Proceedings of the 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), Coimbatore, India, 20–21 April 2018; pp. 368–376.
88. Yao, Z.; Jiang, H.; Cheng, Y.; Jiang, Y.; Ran, B. Integrated schedule and trajectory optimization for connected automated vehicles in a conflict zone. *IEEE Trans. Intell. Transp. Syst.* **2020**, *23*, 1841–1851. [CrossRef]
89. Cao, P.; Xu, Z.; Fan, Q.; Liu, X. Analysing driving efficiency of mandatory lane change decision for autonomous vehicles. *IET Intell. Transp. Syst.* **2019**, *13*, 506–514. [CrossRef]
90. Daniel, A.; Subburathinam, K.; Anand Muthu, B.; Rajkumar, N.; Kadry, S.; Kumar Mahendran, R.; Pandian, S. Procuring cooperative intelligence in autonomous vehicles for object detection through data fusion approach. *IET Intell. Transp. Syst.* **2020**, *14*, 1410–1417. [CrossRef]
91. Heimberger, M.; Horgan, J.; Hughes, C.; McDonald, J.; Yogamani, S. Computer vision in automated parking systems: Design, implementation and challenges. *Image Vis. Comput.* **2017**, *68*, 88–101. [CrossRef]

92. Jeong, Y.; Son, S.; Jeong, E.; Lee, B. An integrated self-diagnosis system for an autonomous vehicle based on an IoT gateway and deep learning. *Appl. Sci.* **2018**, *8*, 1164. [CrossRef]
93. Fényes, D.; Németh, B.; Gáspár, P. A predictive control for autonomous vehicles using big data analysis. *IFAC-Pap.* **2019**, *52*, 191–196. [CrossRef]
94. Fényes, D.; Németh, B.; Gáspár, P. Analysis of autonomous vehicle dynamics based on the big data approach. In Proceedings of the 2018 European Control Conference (ECC), Limassol, Cyprus, 12–15 June 2018; pp. 219–224.
95. Makarova, I.; Buyvol, P.; Gabsalikhova, L.; Pashkevich, A.; Tsybunov, E.; Boyko, A. Improving the reliability of autonomous vehicles in a branded service system using big data. In Proceedings of the 2020 21st International Conference on Research and Education in Mechatronics (REM), Cracow, Poland, 9–11 December 2020; pp. 1–6.
96. Sallab, A.E.; Abdou, M.; Perot, E.; Yogamani, S. Deep reinforcement learning framework for autonomous driving. *arXiv* **2017**, arXiv:1704.02532.
97. Wang, P.; Chan, C.Y. Formulation of deep reinforcement learning architecture toward autonomous driving for on-ramp merge. In Proceedings of the 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC), Yokohama, Japan, 16–19 October 2017; pp. 1–6.
98. Radecki, P.; Campbell, M.; Matzen, K. All weather perception: Joint data association, tracking, and classification for autonomous ground vehicles. *arXiv* **2016**, arXiv:1605.02196 .
99. Zang, D.; Wei, Z.; Bao, M.; Cheng, J.; Zhang, D.; Tang, K.; Li, X. Deep learning–based traffic sign recognition for unmanned autonomous vehicles. *Proc. Inst. Mech. Eng. Part I J. Syst. Control. Eng.* **2018**, *232*, 497–505. [CrossRef]
100. Shaw, D.; Favrat, B.; Elger, B. Automated vehicles, big data and public health. *Med. Health Care Philos.* **2020**, *23*, 35–42. [CrossRef] [PubMed]