ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 8| Issue: 8| August 2022|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2022: 8.205 || ISI Value: 1.188

COMPARATIVE STUDY ON REAL-TIME TRAFFIC STATE ESTIMATION

Akhil Sahukaru

ABSTRACT

When traffic demand exceeds available network capacity, traffic congestion develops.

Lower vehicle speeds, longer journey times, unreliable arrival timings, and lengthier vehicular queueing are all symptoms. Congestion may have a detrimental influence on society by lowering quality of life and increasing pollution, particularly in metropolitan areas. To alleviate traffic congestion, traffic engineers and scientists require high-quality, comprehensive, and precise data to forecast traffic flow. The advantages and disadvantages of various data collecting systems, as well as data attributes such as accuracy, sample frequency, and geographic coverage, vary.

Multisource data fusion improves accuracy and delivers a more complete picture of traffic flow performance on a road network. This study provides a review of the literature on congestion estimation and prediction based on data obtained from numerous sources. An overview of data fusion approaches and congestion indicators that have been employed in the literature to estimate traffic condition and congestion is provided. The outcomes of various strategies are examined, and a disseminative analysis of the benefits and drawbacks of the methods reviewed is offered.

KEYWORDS: *traffic congestion; multi source data fusion; traffic state estimation; data collection*

1. INTRODUCTION

Many Intelligent Transportation Systems (ITS) applications require real-time traffic information, such as incident detection, vehicle routing, traffic signal management, and traffic monitoring. For example, Google has been integrating real-time traffic data with its mapping service since 2007. The data is gathered from a variety of sources, including road sensors, autos, taxi fleets, and, more recently, smartphone users.

However, lack of or incorrect information about the traffic situation might jeopardize drivers' safety, since it is well known that the impacts of congestion on safety are dependent on how shocked drivers are by the congestion. This is especially true if the autonomous road traffic monitoring system believes the road is clear while it is actually crowded, which might result in a rear-end collision owing to the speed disparity between the oncoming vehicle and the anticipated mean traffic speed.

To address these issues, one approach recommended is to rely on systems based on efficient traffic prediction algorithms. The estimation of traffic characteristics (speed, density, and flows) based on a restricted set of traffic variables detected by one or more detectors is known as traffic estimation.

Since the early 1970s, a lot of work has gone into developing methods for predicting road traffic conditions (Knapp, 1972; Nahi and Trivedi, 1973; Grewal and Payne, 1976). The majority of current research on this topic has suggested estimating techniques based on nonlinear Kalman filter extensions (Seo et al., 2017).

The extended Kalman filter (EKF) has been widely employed in traffic estimate systems and over large road traffic networks (Wang and Papageorgiou, 2003, 2005; Wang et al., 2008; Yuan et al., 2014). However, the main disadvantage is the possibility of divergence due to the linearization technique. For road traffic estimates, Pueboobpaphan and Nakatsuji (2006) employed the unscented Kalman filter (UKF) as an alternative to the EKF filter. The performance of the UKF was compared to that of Mihaylova and Boel's particle filter, and the authors concluded that the UKF is a viable approach for traffic flow prediction with a low computing cost (Mihaylova et al., 2006). Hegyi et al. (2006), on the other hand, presented a comparison of several configurations of the UKF filter and its classic competitor, the EKF filter, for freeway traffic state estimation. This comparison is based on state estimation, parameter estimation, joint estimation, and dual estimation performance. The primary results are that the extended Kalman filter and the unscented Kalman filter have virtually identical performance. Work et al. (2008) and Seo et al. (2015) presented the ensemble Kalman filter (EnKF) for predicting the velocity field on a roadway; this filter employs Monte Carlo simulations to address the EKF drawbacks. Thai and Bayen (2015) investigated the difficulty of the EKF and EnKF traffic density estimate filters. With 100 samples, the EKF was determined to be substantially quicker than the EnKF.

In this work, we analyse all of the aforementioned approaches and applications for free real-time traffic status estimate and come up with a research that determines which is more efficient, less expensive, and appropriate for today's reality.

2. PREVIOUS WORK IN THE DOMAIN

2.1 Review Methodology

As previously indicated, all material released in the last twenty to twenty-five years, from 1997 to 2021, was thoroughly investigated. The study was carried out utilising the most frequently utilised scientific databases, which featured research on the review's area of interest.

Key word-based searches were used to locate and filter relevant articles based on a carefully chosen keyword (traffic congestion; multisource data fusion; traffic flow modeling; congestion estimation; traffic state estimation). [3]

The main purpose of the study article is to look at about 30 papers that are thoroughly evaluated and then analysed, with their methods defined. [3]

2.2 Definition of Congestion

Traffic congestion is a condition in which vehicles travel at slower speeds, have longer experience times, and queue for longer periods of time. Since the 1950s, traffic congestion on city avenue networks has increased significantly.

Congestion occurs when traffic demand is high enough that the interaction between cars reduces the pace of the traffic stream.

Extreme traffic congestion occurs when demand exceeds a road's capacity (or the capacity of the junctions along the road). A traffic jam or (informally) a traffic snarl-up occurs when cars are completely halted for long periods of time. Drivers might become upset and engage in road rage as a result of traffic congestion.

Traffic is treated mathematically as a flow past a given point on the route, similar to fluid dynamics.

Some traffic engineers have sought to apply fluid dynamics principles to traffic flow, comparing it to the passage of a fluid through a conduit. Congestion models and real-time observations have revealed that amid heavy yet moving traffic, traffic jams can form spontaneously, prompted by tiny occurrences ("butterfly effects") such as a single motorist's sudden steering action. A circumstance like this is compared to the rapid freezing of supercooled fluid by traffic experts.

Unlike a fluid, however, traffic flow is frequently altered by signals or other occurrences at intersections that disrupt the smooth flow of traffic. Boris Kerner's three-phase traffic theory is an example of an alternative mathematical theory (see also spatiotemporal reconstruction of traffic congestion).

Because theoretical models have a weak association with real observed traffic flows, transportation planners and highway engineers utilise empirical models to anticipate traffic flow. By "platooning" groups of cars and randomising flow patterns inside particular segments of the network, their operational traffic models often employ a combination of macro-, micro-, and mesoscopic characteristics, and may incorporate matrix entropy effects. These models are often calibrated by monitoring real traffic flows on network connections and adjusting the baseline flows accordingly.

A group of MIT mathematicians has created a model that predicts the genesis of "phantom jams," in which little traffic disruptions (such as a motorist slamming on the brakes too hard or going too near to another car) may be magnified into a full-fledged, self-sustaining traffic bottleneck. According to Aslan Kasimov, a lecturer at MIT's Department of Mathematics, "the mathematics of such jams, which the researchers name "jamitons," are startlingly similar to the equations that explain detonation waves produced by explosions." The team was able to solve traffic-jam equations that had been theorised since the 1950s thanks to this discovery.

2.3 Issues Related to Congestion

Many traffic experts and businesses throughout the world deal with traffic congestion and strive to discover various ways to alleviate the situation. Congestion is a major problem in most cities across the world, limiting population movement. In 2019, the top five most congested cities, according to INRIX Research, were (1) Moscow, (2) Istanbul, (3) Bogota, (4)s Mexico City, and (5) So Paulo, where three of the five cities coincide with HERE's rating of the top five most congested cities. The congestion effect rank was established by INRIX Research based on a city's population and the time spent stuck in traffic. [3]

Traffic congestion has a number of negative effects:

- Wasting time of motorists and passengers ("opportunity cost"). As a non-productive activity for most people, congestion reduces regional economic health.
- Delays, which may result in late arrival for employment, meetings, and education, resulting in lost business, disciplinary action or other personal losses.
- Inability to forecast travel time accurately, leading to drivers allocating more time to travel "just in case", and less time on productive activities.
- Wasted fuel increasing air pollution and carbon dioxide emissions owing to increased idling, acceleration and braking.
- Wear and tear on vehicles as a result of idling in traffic and frequent acceleration and braking, leading to more frequent repairs and replacements.
- Stressed and frustrated motorists, encouraging road rage and reduced health of motorists
- Emergencies: blocked traffic may interfere with the passage of emergency vehicles traveling to their destinations where they are urgently needed.
- Spillover effect from congested main arteries to secondary roads and side streets as alternative routes are attempted ('rat running'), which may

ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 8| Issue: 8| August 2022|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2022: 8.205 || ISI Value: 1.188

affect neighborhood amenity and real estate prices.

- Higher chance of collisions due to tight spacing and constant stopping-and-going.

2.4 Data Fusion in Information Technology Sector

Intelligent transportation systems, bioinformatics, cheminformatics, geographic information systems, oceanography, and wireless sensor networks are all examples of where data fusion is used. There are several papers giving a review of DF in ITS. El Faouzi surveyed how DF is utilized in many ITS sectors, including Automatic Terminal Information Service, Automatic Incident Detection, Advanced Driver Assistance, Network Control, Crash Analysis and Prevention, Traffic Demand Estimation, Traffic Forecasting, and Monitoring in [4]. Pattern recognition using adaptive neural networks and clustering methods, as well as identify fusion using Bayesian Decision Theory and Dempster–Schafer evidential reasoning, are among the methods used in the second level to deliver meaningful information from raw data to guide human decision-making. [4] El Faouzi et al. in [5] made a review of the state of practice and prospects for DF in the management of the travel demand which later on came very handy and helpful in future research sectors.

3. TRAFFIC STATE INDICATORS & METHODS

Unique procedures are used by traffic engineers to provide and explain the state of visitors on the road and to estimate the degree of congestion.

The most frequent method is to use traffic flow parameters and Greenshield's fundamental diagram to characterise the status of traffic flow. [6] Another option is to use the Lighthill– Whitham–Richards (LWR) models, which were initially developed in 1955 by Lighthill and Whitham [7], then separately by Richards in 1956. B. Kerner proposed the three-phase traffic theory between 1996 and 2002 [9, 10, 11], which divides congested traffic into two distinct phases: synchronised flow and wide moving jam.

As quantitative congestion indicators, various traffic flow metrics or combinations of parameters can be employed. According to [11], the fundamental traffic characteristics are flow rate q (veh/h), density (veh/km), and speed v (km/h), with the essential relationship q = v. Table 1 lists the other metrics used to define the condition of traffic flow. [6, 12]

The authors of [3, 13] state that delay, density, and Level of Service are the important performance metrics for urban roadways (LOS). LOS is a quality indicator of road network service that is often assessed by speed, density, and volume/capacity ratio [3, 14].

Parameter		Definition	Units
Time Mean Speed	v_t	The arithmetic mean of instantaneous speeds for N vehicles passing an observed road section within a time period *T*	$\frac{m}{s}$
Space Mean Speed	v_s	The arithmetic mean of the instantaneous speeds of the *N* vehicles, which are at the observed road section *d*, at instant time *t* (near zero)	$\frac{m}{s}$
Flow Rate	q	The number of vehicles passing through a given road section, at a given time interval	$\frac{veh}{h}$
Density	ρ	The number of vehicles occupying a given length of a lane or road at an instant time *t* (near zero)	$\frac{veh}{km}$
Throughput	t_v	The throughput is the number of vehicle-kilometers driven for a given length of a road and for a given time period	$\frac{veh \times km}{h}$
Time headway	h_t	Time spacing between the front or back surfaces of the running vehicles in the traffic flow	s
Space headway	h_s	Spatial spacing between the front or back surfaces of the running vehicles in the traffic flow	m
Occupancy	O	The percentage of time at observed road section occupied by vehicles or a total vehicle's dwell time in detection zone at observed interval *T*	%
Queue Length	L_q	Number of vehicles in the queue (intersection, ramp, etc.)	Number of vehicles or km
Travel Time	TT	Time needed for a vehicle to drive from one observed point to another in the traffic network	s

Table 1

To compute the link density and identify the form of the Macroscopic Fundamental Diagram, the authors in [3, 15] coupled flow, which was detected using loop detectors, with journey time, which was measured using GNSS probe taxi cars (MFD).

Other GNSS probe data indications that may be used to determine the condition of traffic flow include Proportion Stopped Time (PST) and Acceleration Noise (AN). The authors of [3, 16, 17, 18] employed a variety of indices to estimate congestion on the connection or network, including the journey time index, space mean speed index, acceleration noise index, buffer index, and planning time index. The survey also mentions indexes for a transport connection and network congestion estimation [3, 13]. A brief summary of congestion indexes is presented in Table 2 [3, 19]. Congestion can also be described using hybrid indicators that combine two or more characteristics.

Parameter		Definition	Equation
Delay	d	The additional travel time experienced by a driver, difference between actual travel time TT and free-flow travel time TT_0	$d = TT - TT_0$
Travel Time Index	TTI	The ratio between delay d and free-flow travel time TT_0	$TTI = \frac{d}{TT_0}$
Speed Reduction Index	SRI	The ratio between free flow speed v_0 and actual speed v difference over free flow speed v_0	$SRI = \frac{v_0 - v}{v_0}$
Buffer Index	BI	The extra time that travelers must add to their average travel time when planning trips to ensure on-time arrival	$BI = \frac{d}{TT_{mean}}$
Travel Rate Index	TRI	The additional time that is required to make a trip because of congested conditions on the roadway.	$TRI = \frac{TT_{peak\ hour}}{TT_{non-peak}}$
Proportion Stopped Time	PST	The ratio of stopped time T_s to the total journey time T_r (running time)	$PST = \frac{T_s}{T_s + T_r}$
Acceleration Noise	AN	Induce fluctuation in speed where Δt_i is the time interval taken for a speed change Δv_i and T_r is vehicle running time	$AN = \sqrt{\frac{1}{T_r} \sum_{i=1}^{N} \frac{\Delta v_i^2}{\Delta t_i}}$
Acceleration Noise Index	ANI	The ratio between actual acceleration noise and acceleration noise in a free flow condition	$ANI = \frac{AN}{AN_0}$

Table 2

4. DATA COLLECTION METHODS AND TECHNOLOGIES

[3, 20] classifies traffic data gathering devices into three categories depending on functionality: point sensors, point-to-point sensors, and area-wide sensors. The term "sensor" here refers to a traffic flow sensor, which is a device or system that may gather data on traffic flow. Inductive loops, piezoelectric sensors, video image sensors, radars, infrared sensors, acoustic sensors, pneumatic road tubes, and magnetic sensors are examples of point sensors. These sensors are used to measure traffic volume, speed, occupancy, and other traffic flow characteristics and are often restricted in spatial coverage [3, 21–23].

Point-to-point sensors, also known as automatic vehicle identification sensors, identify cars at several places throughout the network (AVI). Bluetooth, Wi-Fi, RFID, and Automatic License Plate Recognition are some of the most common technologies used for point-to-point detection (ALPR). These technologies may be used to calculate journey durations, route choice fractions, and origin-destination (O-D) flows [24–26]. Some technologies do not have to fit into one of these categories and can be employed as point or point-to-point sensors. Researchers have used inductive loops for vehicle reidentification and journey time estimate in various articles [3]. As point and point-to-point sensors, cameras and video- and image processing are employed to collect traffic data [3].

Area Wide Sensors are that cover a big area include data collecting systems that allow vehicles to be tracked over a large region. Floating Car Data (FCD) and Cellular Floating Car Data (CFCD) are the most promising (CFCD). FCD data, also known as GNSS probe data, is generated by cellphones or vehicles equipped with GNSS receivers.

ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 8| Issue: 8| August 2022|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2022: 8.205 || ISI Value: 1.188

A detailed study of all the methods and technologies is given below –

	Technologies	Advantages	Disadvantages
Point sensors	Inductive loops	- provide basic traffic parameters (e.g., volume, occupancy, speed, presence, headway) - well-defined detection zone - well-known technology - accurate and reliable traffic data - negligible influence of weather conditions	- installation requires pavement cut and lane closure - spatial coverage is limited - implementation and maintenance costs are high - lifetime depends on pavement quality
	Video detection	- can provide the largest set of data - feasible integration of traffic collection and traffic supervision - can replace several loops - non-intrusive sensor—no pavement cut needed	- affected by weather conditions - calibration issue - cover occurrence
	Radar sensors	- provide speed, vehicle counts, vehicle classification - is not affected by weather conditions - multiple detection zone	- susceptibility to electromagnetic interferences - cover occurrence
	Acoustic sensors	- multiple lane operation available - passive detection - record vehicle's passage, presence, and speed	- spatial coverage is limited - high costs for setting up and maintaining - unsuitable for urban areas with dense traffic

ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 8| Issue: 8| August 2022|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2022: 8.205 || ISI Value: 1.188

	Technologies	Advantages	Disadvantages
	Infrared sensors	- multiple detection zone - small impact of weather conditions	- spatial coverage is limited (depends on sensor type)
	Magnetic sensors	- not affected by weather conditions - can be used where loops are not feasible (e.g., bridge decks)	- spatial coverage is limited
	Piezoelectric sensors	- some models and configurations provide weight in motion and speed	- placed in groove along roadway surface - high costs for setting up and maintaining
Point-to-points sensors	Bluetooth detectors	- can provide travel time, O-D matrices - easy mounting - far greater privacy than ALPR - low energy consumption	- cannot provide volume and vehicle count - low detection accuracy
	Wi-Fi detectors	- easy mounting - suitable for passenger detection - low-cost components	- cannot provide accurate basics traffic parameters
	RFID detectors	- low-cost components - high detection accuracy	- cannot provide volume and vehicle count - small detection zone
	ALPR detectors	- can provide volume, O-D matrices and travel time - high detection accuracy	- privacy issue problematic data protection
Area-wide sensors	FCD	- potential for real-time monitoring - large-scale spatial coverage - location precision is high (10 m) - cost-effective source of data	- limited sample size and time-spatial coverage - high equipment costs
	CFCD	- no additional device is needed - large number of potential probes cost-effective source of data	- for extract data sophisticated algorithms are needed - location precision is low (depends on used location methods and size of mobile network cells) - limited and imprecise spatial coverage
	Airborne imaginary	- mobile multifunctional detection device - can provide density	- limited recording time - affected by weather conditions - high costs
	Social media data	- cheapest data in terms of data availability - potential for real-time data	- low reliability caused by human factor

Table 3

ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 8| Issue: 8| August 2022|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2022: 8.205 || ISI Value: 1.188

5. METHODS USED IN DATA FUSION (FOR PREDICTION, CLASSIFICATION & ESTIMATION)

The authors presented TCE R, a linked matrix and tensor factorization approach for combining multisource data. A method termed search tree-based pattern mining is offered for quickly determining which road segments are likely to face traffic congestion when they are geographically adjacent to each other. Approaches based on recursive Kalman filters give a solution for traffic state estimate and DF. Equations, on the other hand, become computationally costly when data cannot be easily aligned throughout space and time. As a result, the authors present three different DF techniques to handle this challenge, each of which is designed to fuse various traffic sensor data.

Multiple data sources can be fused using the PISCIT, FlowResTD, and Treiber–Helbing filter (EGTF), as long as it is reasonable to determine under which traffic circumstances each of these was gathered (congested or free flow). [3]

5.1. Statistic Methods

Research was conducted on how to use social media as an auxiliary data source and combine it with GNSS probe data to improve traffic congestion estimation. The writers gathered a large number of tweets on various traffic occurrences, such as traffic jams, accidents, and road construction. The researchers then suggested an enhanced Coupled Hidden Markov Model that can successfully integrate GNSS probe data with traffic-related tweets to better precisely estimate traffic conditions in an arterial network. In compared to earlier methodologies, the experimental findings showed that the model performed better.

Zhu et al. [27] integrated data from three sources: bus-based GPS data, inductive loop detector data, and mobile phone network data, using three distinct DF algorithms. To fuse different data resources and provide more accurate trip times, the hybrid technique surpasses the weighted mean approach and artificial neural networks. The findings show that combining several data sources does not always improve the accuracy of journey time prediction. The accuracy of separate data sources affects travel time estimation. When highly connected data sources are combined, the consequence might be disastrous. The findings also indicate that, even in densely populated locations, GPS data paired with inductive loop detector data may offer accurate trip time estimates for broad traffic streams under a variety of traffic conditions. [3]

5.1. Deep Neural Network

An examination of the impact of meteorological conditions on traffic speed in metropolitan areas was conducted. The authors employed the Long Short-Term Memory Neural Network to do this (LSTM-NN). LSTM-Appl. Sci. 2021, 11, 2306 13 is capable of "forgetting" or storing knowledge for a longer length of time. Because of this characteristic, LSTM-NN models outperformed SVM, Kalman Filter, and ARIMA models in forecasting speed. The researchers used data from inductive loops as well as meteorological data such as rainfall and temperature. They used an urban arterial route in Greater Manchester for the testing scenario, and the model, which included meteorological data with inductive loops, produced the best forecast results in terms of lowest absolute error. ARIMA was surpassed by this approach by several orders of magnitude. [3]

Then there was a proposal for Deep Ensemble layered Long Short-Term Memory, a deep learning-based framework that integrated road network, weather, and traffic data to anticipate long-term traffic time (DE-SLSTM). They use the "cost sensitive" technique in the suggested framework to increase forecast accuracy during rush hours due to the difficulty of estimating traffic time during congestion. The suggested framework performs well and matches the ground truth better than Google maps.

There is a detailed study of all the methods in Data Fusion – (Representation of Methods for Data Fusion (DF) (MF, TF–Matrix and Tensor Factorization; STAT—Statistical; ANN—Artificial Neural Network; MM—Markov Model; KF—Kalman Filter; IP—Image Processing; DNN—Deep Neural Network; CLUS—Clustering; OPT—Optimization; FUZ—Fuzzy; CLA—Classification).

ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 8| Issue: 8| August 2022|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2022: 8.205 || ISI Value: 1.188

	TF	MF	STAT	ANN	MM	KF	IP	DNN	CLUS	OPT	FUZ	CLA
Wang et al. [71,97]	X 2017	X 2016										
Zhu et al. [91]			X 2018	X 2018								
Wang et al. [79]			X 2016		X 2016							
Ji et al. [37], Ambühl et al. [82], Bhaskar [83], Li et al. [85], Havyarimana et al. [101]			X 2018, 2016, 2010, 2016, 2020									
Patire et al. [80], Kong et al. [72]			X 2015, 2009			X 2015, 2009						
Jiang et al. [81]			X 2017				X 2017					
Essien et al. [75], Yuan et al. [46], Chou et al. [95], Rodrigues et al. [96], Gu et al. [102], Liang et al. [103]								X 2019				
Zheng et al. [84]			X 2018						X 2018			
Wu et al. [78]										X 2015		
Choi et al. [92]			X 2002								X 2002	
Sohn et al. [91]			X 2003	X 2003					X 2003			X
Mil et al. [93]			X 2018									X 2018
Li et al. [96]			X 2019				X 2019					
Luo et al. [100], Guo et al. [104]			X 2019					X 2019	X 2019			
Ke et al. [97], Hu et al. [98]							X 2019	X 2019				
Guan et al. [99]								X 2019				X 2019
Usage of method	2, 4%	2, 4%	17, 34%	2, 4%	1, 2%	2, 2%	4, 8%	11, 22%	4, 8%	1, 2%	1, 2%	3, 6%

Table 4

ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 8| Issue: 8| August 2022|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2022: 8.205 || ISI Value: 1.188

6. YOUR CONTRIBUTION

- A comprehensive literature review was conducted using a keyword-based search of academic research databases and the systematic selection of highly relevant publications from the search results based on their influence on the scientific community.
- Data fusion from numerous sensors is used to cover traffic congestion estimate studies in metropolitan networks.
- Analyze the goals of congestion estimation and data fusion, such as increasing efficiency or accuracy, as well as the various data fusion approaches employed and their performance.

7. CONCLUSION

Because one knowledge collection technique can contribute traffic data while others are absent, inaccurate, or ineffective, research presented throughout this study demonstrates that utilising multisource DF will boost estimation reliability and hardiness.

Several prospective research avenues to overcome the shortcomings of existing approaches were highlighted as a result of the survey conducted on the selected studies. Two dominant approaches in estimation of congestion that were mostly used lately are statistical and deep learning methods:

- When dealing with complicated and highly nonlinear data, statistical approaches can give insights into traffic flow conditions, but they fail when dealing with complex and highly nonlinear data. A statistical approach is utilised to provide insights into the data's linkages and structure, or to construct a model that may forecast future traffic conditions. Statistical approaches have more "understandable" mathematical underpinnings than certain deep learning methods since they are based on strong and widely acknowledged mathematical foundations.
- Deep learning algorithms, which have been widely used, develop "intelligent" models that employ a significant quantity of data to get meaningful insights about traffic flow and recognise distinct patterns. Although deep learning is more flexible than statistics, there isn't necessarily a mathematical reason for why one method performs better than another. There are two techniques to employing deep neural network (DNN) methods: (i) a mix of image processing-related approaches that make use of convolutional neural networks, and (ii) time-series analysis that makes use of the long-short term memory network. DNNs have been widely used to solve a variety of transportation issues, owing to the fact that they are extremely versatile, accurate, and convenient mathematical models that can readily replicate numerical model components. Because of their capacity to cope with huge volumes of multidimensional and multisource data, they have mostly been employed as a data analysis approach.
- DNN approaches are more flexible than statistical methods since the functional form is approximated through learning rather than assumed a priori as in statistics. DNN-based models, on the other hand, might be computationally and memory intensive.

Statistical analysis is the most prominent data fusion strategy from 2010 to 2020, while DNNs are the most dominant in recent approaches. It may be inferred that, in general, traffic flow prediction should be done utilizing data from numerous diverse sources to avoid biases introduced by particular data gathering systems. A standardized testing dataset from multiple multisource data, which would provide real numerical proof of how successful a technique is, given the available data, would be another useful tool in the development of prediction approaches, such as data mining use.

7.1 Future Work / Potential Research

This study did not have non-CV data (Connected Vehicle), such as data from social media and/or news feed.

Machine learning algorithms and recommendation engines using social media data from sites like twitter and fakebook, can be incorporating data for traffic estimation.

Data from news or audio signals can also be incorporated and be taken into consideration.

[28] The following directions should be considered in future data homogeneity analysis research:

- Expanding the correlation analysis selection range by increasing the data categories used for missing traffic volume estimation;
- Investigating some correlation analysis methods that take into account both data distribution variance and time fluctuation differences;
- Developing reasonable correlation judgement criteria based on different missing volume scenarios. [28]

8. REFERENCES

1. *Asmâa Ouessai, Mokhtar Keche, REAL-TIME FREEWAY TRAFFIC STATE ESTIMATION BASED ON THE SECOND-ORDER DIVIDED DIFFERENCE KALMAN FILTER, Transport and Telecommunication, 2019, volume 20, no. 2, 114–122*
2. *Ren Wang, Daniel B. Work (PI), Joint parameter and state estimation algorithms for real-time traffic monitoring, NEXTRANS Project No 097IY04.*
3. *Dominik Cvetek, Mario Muštra, Niko Jelušić and Leo Tišljarić, "A Survey of Methods and Technologies for Congestion Estimation Based on Multisource Data Fusion", Multidisciplinary Digital Publishing Institute, Appl. Sci. 2021, 11(5), 2306*
4. *El Faouzi, N.-E.; Leung, H.; Kurian, A. Data fusion in intelligent transportation systems: Progress and challenges—A survey. Inf. Fusion 2011, 12, 4–10*
5. *Zegras, C.; Pereira, F.; Amey, A.; Veloso, M.; Liu, L.; Bento, C.; Biderman, A. Data Fusion for Travel Demand Management: State of the Practice and Prospects. In*

ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 8| Issue: 8| August 2022|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2022: 8.205 || ISI Value: 1.188

Proceedings of the TDM'08, Travel Demand Management Symposium, Arlington, VA, USA, 15–16 November 2008; pp. 1–17.
6. *May, A.D. Traffic Flow Fundamentals; Prentice Hall: Upper Saddle River, NJ, USA, 1990.*
7. *Lighthill, M.J.; Whitham, G.B. On kinematic waves I. Flood movement in long rivers. Proc. R. Soc. London. Ser. A Math. Phys. Sci. 1955, 229, 281–316.*
8. *Richards, P.I. Shock Waves on the Highway. Oper. Res. 1956, 4, 42–51.*
9. *Kerner, B.S. Experimental Features of Self-Organization in Traffic Flow. Phys. Rev. Lett. 1998, 81, 3797–3800.*
10. *Kerner, B.S. The Physics of Traffic. Underst. Complex Syst. 2004, 12, 25–30.*
11. *Kerner, B.S. Congested Traffic Flow: Observations and Theory. Transp. Res. Rec. J. Transp. Res. Board 1999, 1678, 160–167.*
12. *Geroliminis, N.; Daganzo, C.F. Macroscopic modeling of traffic in cities. In Proceedings of the TRB 86th Annual Meeting, Washington, DC, USA, 21–25 January 2007; pp. 7–413.*
13. *Mohan Rao, A.; Ramachandra Rao, K. Measuring Urban Traffic Congestion—A Review. Int. J. Traffic Transp. Eng. 2012, 2, 286–305.*
14. *Highway Research Board. Highway Capacity Manual; Transportation Research Board: Washington, DC, USA, 2000.*
15. *Ji, Y.; Xu, M.; Li, J.; Van Zuylen, H.J. Determining the Macroscopic Fundamental Diagram from Mixed and Partial Traffic Data. Promet—TrafficTransp. 2018, 30, 267–279.*
16. *Carli, R.; Dotoli, M.; Epicoco, N. Monitoring traffic congestion in urban areas through probe vehicles: A case study analysis. Internet Technol. Lett. 2018, 1, e5.*
17. *Tahmasseby, S. Traffic Data: Bluetooth Sensors vs. Crowdsourcing—A Comparative Study to Calculate Travel Time Reliability in Calgary, Alberta, Canada. J. Traffic Transp. Eng. 2015, 3.*
18. *Stipancic, J.; Miranda-Moreno, L.; Labbe, A. Measuring Congestion Using Large-Scale Smartphone-Collected GPS Data in an Urban Road Network. In Proceedings of the Conference and Exhibition of the Transportation Association of Canada, Toronto, ON, Candada, 25–28 September 2016.*
19. *Toledo, C.A.M. Congestion Indicators and Congestion Impacts: A Study on the Relevance of Area-wide Indicators. Procedia—Soc. Behav. Sci. 2011, 16, 781–791.*
20. *Antoniou, C.; Balakrishna, R.; Koutsopoulos, H.N. A Synthesis of emerging data collection technologies and their impact on traffic management applications. Eur. Transp. Res. Rev. 2011, 3, 139–148. [CrossRef]*
21. *Hazelton, M. Estimating vehicle speed from traffic count and occupancy data. J. Data Sci. 2004, 2, 231–244*
22. *Bugdol, B.; Segiet, Z.; Kre¸cichwost, M.; Kasperek, P. Vehicle detection system using magnetic sensors. Transp. Probl. 2014, 9, 49–60*
23. *Lopez, A.A.; De Quevedo, A.D.; Yuste, F.S.; DeKamp, J.M.; Mequiades, V.A.; Cortes, V.M.; Cobena, D.G.; Pulido, D.M.; Urzaiz, F.I.; Menoyo, J.G. Coherent Signal Processing for Traffic Flow Measuring Radar Sensor. IEEE Sens. J. 2018, 18, 4803–4813. [CrossRef]*
24. *Araghi, B.N.; Krishnan, R.; Lahrmann, H. Mode-Specific Travel Time Estimation Using Bluetooth Technology. J. Intell. Transp. Syst. 2015, 20, 219–228. [CrossRef]*
25. *Yuan, J.; Yu, C.; Wang, L.; Ma, W. Driver Back-Tracing Based on Automated Vehicle Identification Data. Transp. Res. Rec. J. Transp. Res. Board 2019, 2673, 84–93. [CrossRef]*
26. *Barceló, J.; Montero, L.; Bullejos, M.; Serch, O.; Carmona, C. A Kalman Filter Approach for Exploiting Bluetooth Traffic Data When Estimating Time-Dependent OD Matrices. J. Intell. Transp. Syst. 2012, 17, 123–141. [CrossRef]*
27. *Zhu, L.; Guo, F.; Polak, J.W.; Krishnan, R. Urban link travel time estimation using traffic states-based data fusion. IET Intell. Transp. Syst. 2018, 12, 651–663. [CrossRef]*
28. *JipingXing WeiWu QixiuCheng RonghuiLiu Traffic state estimation of urban road networks by multi-source data fusion: Review and new insights, Physica A: Statistical Mechanics and its Applications Volume 595, 1 June 2022, 127079*