IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 26, 2021, accepted March 17, 2021, date of publication March 30, 2021, date of current version April 8, 2021.

*Digital Object Identifier 10.1109/ACCESS.2021.3069770*

# Review of Data Fusion Methods for Real-Time and Multi-Sensor Traffic Flow Analysis

**SHAFIZA ARIFFIN KASHINATH[1,2], SALAMA A. MOSTAFA[1], AIDA MUSTAPHA[1], (Member, IEEE), HAIRULNIZAM MAHDIN[1], DAVID LIM[2], MOAMIN A. MAHMOUD[3], MAZIN ABED MOHAMMED[4], BANDER ALI SALEH AL-RIMY[5], MOHD FARHAN MD FUDZEE[1], (Senior Member, IEEE), AND TAN JHON YANG[2]**

[1]Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia, Batu Pahat 86400, Malaysia
[2]Engineering Research and Development Department, Sena Traffic Systems Sdn. Bhd., Kuala Lumpur 57000, Malaysia
[3]College of Computer Science and Informatics, Universiti Tenaga Nasional, Kajang 43000, Malaysia
[4]College of Computer Science and Information Technology, University of Anbar, Ramadi 31001, Iraq
[5]Faculty of Engineering, Universiti Teknologi Malaysia, Johor Bahru 81310, Malaysia

Corresponding author: Salama A. Mostafa (salama@uthm.edu.my)

This project is funded by the Ministry of Higher Education Malaysia under the Malaysian Technical University Network (MTUN) grant scheme Vote K235 and SENA Traffic Systems Sdn. Bhd.

**ABSTRACT** Recently, development in intelligent transportation systems (ITS) requires the input of various kinds of data in real-time and from multiple sources, which imposes additional research and application challenges. Ongoing studies on Data Fusion (DF) have produced significant improvement in ITS and manifested an enormous impact on its growth. This paper reviews the implementation of DF methods in ITS to facilitate traffic flow analysis (TFA) and solutions that entail the prediction of various traffic variables such as driving behavior, travel time, speed, density, incident, and traffic flow. It attempts to identify and discuss real-time and multi-sensor data sources that are used for various traffic domains, including road/highway management, traffic states estimation, and traffic controller optimization. Moreover, it attempts to associate abstractions of data level fusion, feature level fusion, and decision level fusion on DF methods to better understand the role of DF in TFA and ITS. Consequently, the main objective of this paper is to review DF methods used for real-time and multi-sensor (heterogeneous) TFA studies. The review outcomes are (i) a guideline of constructing DF methods which involve preprocessing, filtering, decision, and evaluation as core steps, (ii) a description of the recent DF algorithms or methods that adopt real-time and multi-sensor sources data and the impact of these data sources on the improvement of TFA, (iii) an examination of the testing and evaluation methodologies and the popular datasets and (iv) an identification of several research gaps, some current challenges, and new research trends.

**INDEX TERMS** Intelligent transportation systems, traffic flow analysis, data fusion, real-time processing, multi-sensor, heterogeneous data, machine learning.

## I. INTRODUCTION

Data has become a central and dominant element in every decision-making phase as a consequence of technological advancements and demands from different disciplines. When dealing with real-time decision-making systems, it becomes critical, and it has already benefited a variety of fields, including transportation [1]–[3], environmental [4], health care [5], smart card [6], image processing [7], structures [8], traffic flow analysis [9]–[17], Internet of Things (IoT) [18] and Big Data [9], just to name a few.

The availability of cloud computing and IoT strengthen the infrastructure of transportation systems and stimulate the development of ITS globally [19]. One of the contributions that the ITS has attempted to make is a long-term increase in traffic research. TFA is a foundational analysis that will help organizational traffic decisions and future improvement planning. Reliable traffic data collection generates traffic insights within a pre-defined period, either from a real-time or multi-sensor system environment. The ITS performs better interpretation of the observed traffic conditions by

The associate editor coordinating the review of this manuscript and approving it for publication was Sotirios Goudos.

considering various data sources across different data providers. Among the data providers are cameras, global positioning systems (GPS)s, probe vehicles, social media, radars, and loop sensors. Their desired integration into a specific ITS model contributes to improving the ITS [17], [20]. Engaging various field data collection methods reduces the chances of uncertainties of only relying on individual data sources [21].

To compose this paper, we extract research articles and other related literature papers. We construct meaningful search keywords for effective findings, including DF review, DF techniques, DF traffic, traffic state estimation, DF applications, DF framework, DF intelligent transportation system, and DF evaluation technique. We use several academic databases, including Springer Link, IEEE Xplore, and ScienceDirect. Figure 1 shows the number of research and review papers discussed in this review from 2010 until 2020, whereas the number on top of the bar shows the number of DF review papers written in that year.

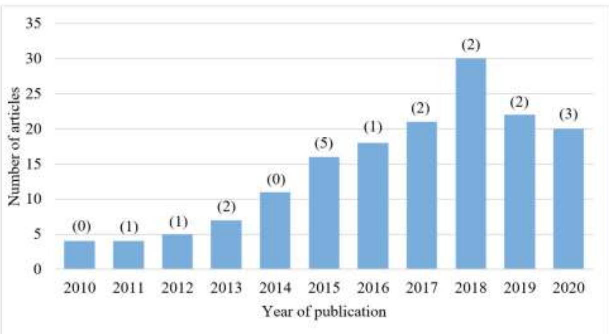



**FIGURE 1. The reviewed papers on DF and TFA from 2010 to 2020.**

As seen in Figure 1, there are 19 papers that focus on DF review writing between these years. Most of the review study discuss about general DF ideas, techniques and models [18], [22]–[29]. Quite a smaller number of literature papers focus on a specific domain of study. This paper offers TFA as a domain study of the DF review. It focuses on studying various traffic conditions, including travel time, traffic speed, traffic density, lane changing behavior, traffic congestion, and traffic incident prediction. The review considers both real-time traffic data and various sensors that are fixed to the real traffic environments of ITSs. This discussion emphasizes the relation between the TFA models with the mechanisms, algorithms, techniques, and methods that construct these models for performing preprocessing, filtering, estimation, forecasting, evaluation, and behavior identification.

The paper's contents are organized as follows: Section I is the introduction, followed by Section II, focusing on traffic sensing methods and traffic variables. Discussion of DF is elaborated in Section III. Section IV emphasizes DF implementation in TFA and contribution to the domain field and providing a few of the most frequently used DF techniques in TFA study. Section V focuses on analysis and discussion of DF-TFA research and applications challenges. Finally, Section VI presents the conclusion and suggests several recommendations.

## II. TRAFFIC SENSING METHODS AND VARIABLES

Different statistical standard variables are used to measure scores on different scales based on sensing methods and data processing techniques (e.g., maximum, minimum, mean, and standard deviation). In TFA, there are a plethora of variables produced, such as speed, vehicle count, vehicle presence, and location. Sensing and gathering information steps are the most fundamental aspect of better control of any management and monitoring systems [30]. Sensing methods are crucial in traffic management systems to measure and evaluate roads' efficiencies [31]. Different traffic sensing devices give various signals or indicators based on their ability and designed functionality [32], [33].

### A. TFA ATTRIBUTES

TFA discusses all sorts of traffic measurement, estimation, and prediction contributions to various areas of ITS and traffic management. Table 1 presents some relationships between traffic state attributes within the related TFA studies.

**TABLE 1. Traffic state.**

| Traffic state | TFA Studies | Ref. |
|---|---|---|
| Vehicle volume, speed, occupancy | Travel time estimation | [34] [35] [36] |
| Vehicle volume, GPS | Traffic speed estimation | [37] |
| Vehicle volume, GPS | Traffic speed prediction | [38] |
| Vehicle volume, speed, location, weather, wind speed | Traffic flow forecasting | [32] [39] [40] [41] [42] |
| Fusing point and zone-based data | Travel time and density estimation | [43] [44] |
| Steering wheel, speed | Lane changing behavior | [45] |
| Flow, travel time | Traffic density estimation | [43] [46] [47] |
| GPS, flow | Traffic congestion prediction | [9] [48] |
| Flow, speed, occupancy | Traffic incident prediction | [49] [50] |

Various traffic variables are reflected by different traffic states of traffic flow parameters. For example, the traffic flow parameter of vehicle volume represents the total number of vehicles on the road as observed over a period of time [34]. Similarly, speed denotes the distance a vehicle traveled in time units, while occupancy represents the extent of the road the vehicles occupy [35]. While vehicles' locations are provided by the GPS [37], the weather and wind speed indicate the conditions that may occur in the area.

In other words, estimation of traffic conditions is determined by the current states of traffic parameters as indicated by the collected data, for example, estimating the average speed from multiple data sources [37]. By the same token, research in predicting traffic conditions yields prediction results by modeling to determine the future situation. For example, Li *et al.* [38] suggest a hybrid method based on deep feature fusion modeling to achieve speed prediction. Traffic density is estimated to evaluate a road's density, such as an urban signalized junction, by integrating flow and travel time state attributes [43]. Traffic flow is the most fundamental criterion of understanding road capacity and traffic congestion, and it is divided into long-term and short-term predictions [42]. It is an essential measurement for travel navigation decisions [40], transportation management [41], smart city planning [42], and others. Most of the research conducted in this specific area aims to propose a better traffic handling mechanism by making full use of the various source of data, such as GPS, the incoming flow of vehicle, the outgoing flow of vehicle, even meteorological data, including weather, temperature, and wind speed [40], [41]. Traffic congestion highly relies on GPS and sensor data [9].

Adetiloye and Awasthi [9] categorize congestion as (1) high congestion prediction, (2) possible congestion, (3) medium congestion prediction, and (4) low congestion prediction based on the study conducted. Response to the pre-defined variable state in this study contributes to traffic management decision-making procedure. Travel time estimation requires accurate vehicle count data from loop detectors and vehicles' locations (from GPS), whether it is measured from road traffic or freeways [34]–[36]. Speed estimation could be generated with the availability of vehicle volume data and GPS collected from vehicles, phones, or navigator devices [38], [51].

A traffic incident is another area of traffic studies that predicts road incidents based on traffic behavior by considering a few traffic state attributes, such as vehicle flow, speed, and occupancy [51]. A road incident is closely related to the severity of congestion. It is useful for the emergency response team to facilitate efficient road traffic management to avoid prolonged road congestion [50]. Lane changing is an everyday event among road users in all circumstances, and this area requires an initial knowledge of classifying and differentiating driving behavior based on changing state of the object or vehicle being observed, in fact, data such as steering wheel angle, accelerations, and vehicle speeds are required [45].

### B. TRAFFIC SENSING METHODS

There are few technologies or devices that are used for traffic data gathering, including, Global Positioning System (GPS) equipped in vehicles [13], [31], wireless communications [31], loop sensor or fixed detector [17], [32], [33], [52], Remote Traffic Microwave Sensors (RTMS) [13], radar [53], Automatic Vehicle Identification (AVI) Bluetooth [54] and many more. Integration of different sensor technologies could bring a convincing result to achieve ITSs [55].

Single sensor technology is not an all-time practical mechanism to illustrate a complete cognizance of a domain problem, while accuracy level and data certainty might be questionable. Having an incomplete and inaccurate data collection phase may affect estimation or forecast results [3], [34]. In ITS, the needs of multiple sensors have become increasingly crucial for multiple purposes, such as (1) lane management, (2) surveillance, (3) parking management, (4) automatic tolling, (5) special event transportation, (6) intersection management and a lot more. There are few types of sensor technologies available to gather traffic data, and each of them has its strength and weakness, as shown in Table 2.

**TABLE 2.** Category of sensors and variables.

| Sensor/Method | Description | Variables |
|---|---|---|
| Inductive Loop Detector | Detects vehicles' movement, presence, count, and occupancy. Reliable under various weather conditions | Vehicle count, vehicle presence |
| Magnetic Sensors | Detects vehicle's presence, identify stopped and moving vehicles | Vehicle presence |
| Camera | Detects vehicles across several lanes, vehicle classification, flow rate, occupancy, and speed. Cameras are linked to a computer with an intelligent algorithm to retrieve traffic parameters. It is of low cost, easy to install and maintain. | Flow rate, occupancy, speed, density, queue length |
| Radar | Uses radio waves to detect vehicles, measure speed, and detect movement direction. It is of high cost, difficult to install, and difficult to maintain. | Vehicle count, speed, direction |
| Infrared | Detects infrared radiation through sender and receiver parts. It can measure speed, vehicle volume, and lane occupancy. It is of a low cost but difficult to be maintained. | Speed, vehicle count, occupancy |
| Ultrasonic | Uses ultrasonic waves to detect vehicle presence and occupancy A low cost but difficult to maintain sensor. | Vehicle count, vehicle presence, occupancy |
| Radio Frequency Identification (RFID) | Uses electromagnetic fields to identify and track vehicles. It is mainly used for toll management. | Target identification |
| GPS | Uses Satellite-based sensing to provide information on vehicle location. It is of a high cost, difficult to install, and difficult to maintain. | Coordinate, count, speed, direction |
| RTMS | Uses radar technology for vehicle detection | Average vehicle length, Speed |
| AVI Bluetooth | Detects traffic stream by continuous Bluetooth sensor | Travel time |
| Cellphone-probe | Provides geographic location | Cellphone count, pseudo speed |

Sensors create a raw input signal to the system in various formats, depending on the type of data being collected [56]. The input signals from sensors produce data, filtered into meaningful features before any decision is made [57], [58]. The collective input from two or more sensors could lead to a better traffic management system [33]. Zhang *et al.* [57] generate real-time traffic state estimation by merging data from a loop sensor and GPS. Deng *et al.* [54] incorporate vehicle detectors, automatic vehicle identification (AVI) Bluetooth, and GPS as their sensing device to determine traffic estimation. Loop detectors and GPS data are two input sources to Jiang *et al.* [17] research on urban expressways study.

Subsequently, multimodal sensing tends to give a better scenario description to an existing problem [59]. It contributes to big data growth in general. Multimodal sensing ensures input robustness and trust, reduces the risk of missing data, and increases input accuracy when cross-validation is performed during the fusion process [60]. For multiple sensors that are sending data to a system, they are not necessarily installed side by side, and they might measure different criteria of the input. This issue requires applying some data association and estimation [58].

## III. DATA FUSION IN TFA

This section provides a general idea of DF processes, data properties, and their implementation in TFA. There are several main objectives of applying DF methods in TFA. DF methods help to overcome inconsistency and imperfection during data collection. DF methods extract higher-level information from raw data sources that reduce data capacity and improve its quality. DF methods improve data completeness and reliability by enabling systems to gather data from multiple sources and with different properties.

### A. DATA PROPERTIES

In any estimation or prediction study, the more consistent data collected, the better the accuracy of results produced [39]. Rich features in datasets require refinement and exploration to maximize the output to the field. A multi-sensor offers diverse and heterogeneous data structures with high variability of formats. Heterogeneous forms of data primarily from various sources can compensate for incomplete data from sensors [61]. It is a critical task of DF preprocessing to eliminate any possibility of having low quality, redundant or ambiguous data in the initial stage. Real-time data is mostly collected from sensors at a specific interval, which requires a particular process to collect, control, and monitor the resources [62]. DF on multi-features data are not only applicable to a single domain but can be blended among different domains [63].

Correlating heterogeneous data is significant in real-time system environments to accurately describe real-world scenarios [64]. A variety of data characteristics proves its capability to describe a problem being observed from different angles [37], [74]. Integrating data from multiple sources increases the results' accuracy and stability [70]. Multi-sensor, heterogeneous, real-time, or combination of those features would stimulate a DF technique's suitability for further implementation. For instance, Ning *et al.* [70] demonstrate a model incorporating data from multi-sensor and heterogeneous in the real-time system environment, such as loop detectors and probe vehicles. They found that integrating cameras and mobile phones can achieve comprehensive traffic state estimation. Cipriani *et al.* [71], Mil and Piantanakulchai [72], and Jiang *et al.* [17] integrate multi-sensor and heterogeneous data, including loop detectors and probe vehicles, to come out with their state estimation. Deng *et al.* [54] propose fusing data from loop detectors, AVI Bluetooth, and GPS to improve traffic state estimation on freeways. Examples of TFA data characteristics are shown in Table 3.

**TABLE 3. Characteristics of traffic flow datasets.**

| Traffic flow study | Sensors | Multi-sensor | Hetero-generous | Real-time | Ref. |
|---|---|---|---|---|---|
| Traffic state estimation | Loop detectors, Bluetooth, GPS | √ | √ | √ | [15] |
| Travel time estimation | Probe vehicles | - | √ | √ | [16] |
| Traffic state and emission estimation | Loop detectors, probe vehicles | √ | √ | √ | [17] |
| Traffic state estimation | Probe vehicles | - | √ | √ | [65] |
| Traffic state estimation | License plate, vehicle count, flow rate, speed | √ | √ | √ | [66] |
| Traffic state estimation | Loop detectors, AVI Bluetooth, GPS | √ | √ | √ | [54] |
| Travel time estimation | Inductive loop, dedicated short-range communications | √ | √ | √ | [67] |
| Traffic state estimation | Loop detectors, connected vehicles | √ | √ | √ | [68] |
| ITS computational data | SCATS loop detectors and probe vehicles | √ | √ | √ | [69] |
| Traffic state estimation | Loop detectors, probe vehicles | √ | √ | √ | [70] |
| Traffic state estimation | Loop detectors, probe vehicles | √ | √ | √ | [71] |
| Travel time estimation | Loop detectors, probe vehicles | √ | √ | √ | [72] |
| Traffic state estimation | Small imaging satellite, connected vehicles | √ | √ | √ | [73] |
| Travel time estimation | License plate recognition, Geomagnetic detector data, Floating car data | √ | √ | √ | [74] |
| Travel time estimation | Vehicle Mass and Road Grade | √ | √ | √ | [75] |
| Traffic state estimation | Connected vehicles | - | √ | √ | [76] |
| Travel time estimation | GPS, inductive loop sensors, mobile phone network | √ | √ | - | [77] |
| Traffic flow prediction | GPS | - | √ | √ | [78] |
| Traffic flow forecasting | Vehicle flow | - | - | √ | [79] |

Similarly, to introduce a robust traffic estimation method, Nantes *et al.* [15] use data from loop detectors, Bluetooth, and GPS by implementing a heterogeneous data source. They conclude that the greater the amount of data being fed, the higher the accuracy of the outcome. Seo and Kusakabe [73] combine small imaging satellites and connected vehicles to generate traffic state estimation, which does not require any parameter calibration on any device input. Guo and Yang [74] propose a robust travel-time estimation method based on license plate recognition, geomagnetic detector data, and floating car data as traffic data input. Zhang *et al.* [75] perform real-time estimation of road infrastructure grading based on data collected from GPS, an inertial navigation system, and wheel-speed. Khan *et al.* [76] work on a real-time traffic estimation study by integrating connected vehicle technology with the multi-sensors approach. Tak *et al.* [35] merge traffic data collected from an inductive loop and dedicated short-range communications to produce a real-time prediction of travel time.

### B. DATA FUSION APPROACH IN TFA

DF methods reconstruct data from different data sources to find new correlations and combinations that generate new data and produce better decisions and actions [51]. Formulating DF methods or algorithms consists of four basic processes: preprocessing, filtering, decision, and evaluation [27]. Achieving high-quality results from a DF model depends on the algorithm combinations choice to perform these processes, the quality of the inputs, and the type of outputs. These steps attempt to combine data to reveal meaningful features that achieve higher accuracy values and assist in improving the decision-making process. Figure 2 shows a general DF model of multi-sensor heterogeneous data.

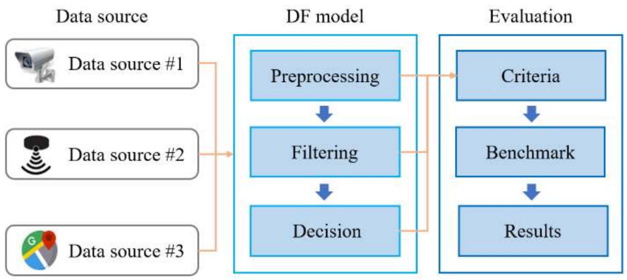



**FIGURE 2.** General DF model.

On the other hand, for heterogeneous and complex multimodal sensing, DF can be divided into three abstraction levels: (1) data level fusion, (2) feature level fusion, and (3) decision level fusion [80]–[84]. Figure 3 summarizes the taxonomy of DF traffic parameters/variables in the TFA studies. DF systems need to handle a heavy communication load to produce reliable and accurate results [80]. Data level fusion manifests high complexity, especially for incoming data with various characteristics. A data level fusion or preprocessing method operates on the collected data in its original form. It entails combining, associating, and formatting the data to prepare for raw data [85]. This level implements various techniques, including noise removal, outliers, sudden spike [64], [70].

Wang *et al.* [12] improve data quality by performing data cleaning and denoising using the Kalman Filter (KF) technique. For state estimation studies, it is common to prepare the data in batches. Pamuła and Król [79] perform discretization to the dataset to group the data based on a specific interval to be inputted to the proposed model. In contrast, Liu *et al.* [66] perform a weighted average to determine vehicle speed for an interval of two minutes. The missing value is another challenge when dealing with datasets, and to a certain field of study, this condition produces uncertainty during model evaluation. An *et al.* [40] handle this situation by performing estimation at the data level to avoid having less sampling data from the dataset by implementing a proximity alternative method.

Feature level fusion is a process of merging and filtering data to extract features from different sources or sensors to achieve meaningful features and comprise statistical, segmentation, or clustering process [70], [84]. A combination of features entails making a more realistic condition. Liu *et al.* [86] use speed and density features to come out with freeway state estimation. Mehrannia *et al.* [49] combine features from different sensor systems for traffic incident detection. Adetiloye and Awasthi [9] engage with the heterogeneous and heterogeneous data types in the proposed model. Features from different sources are used to make congestion predictions. Zhu *et al.* [68] propose a DF model that fuses loop detector and mobile phone data to estimate travel time. Other than numeric data, the filtering process can also be done for motion type of data input. Gao *et al.* [45] combine features from video data with lane boundary and distance to determine lane changing of vehicles on the road.

Decision level fusion is the stage where the system's decision is made by selection, inferencing, and reasoning [87]. It comprises the classification, prediction, or estimation process [71]. Soua *et al.* [88] fuse several traffic flow decisions from event-based data streams to predict the final traffic flow. Nae and Dumitrache [89] produce a final decision by synthesizing a time-based system and sensor-based system as a mechanism to manage traffic light timing strategies. An *et al.* [40] fuse decisions from three systems: accident information, priority vehicle transition, and road information to be used by traffic signal control that handles both coordinated and isolated intersections.

The evaluation process finds the correlation between estimated or found results with the actual scenario or benchmarking data. It can be done quantitatively or qualitatively, depending on the modeling of the DF methods [64]. The evaluation process helps to describe the efficiency and effectiveness of the implemented DF model or method. These issues are discussed in detail in the next section.

## IV. MODELING DATA FUSION

### A. DF METHODS

Each DF technique, algorithm, or method is strengthened by adopting a specific algorithm and/or mathematical model to solve the domain's particular problem. This section discusses a few popular DF algorithms, methods, and techniques in building DF models.

#### 1) KALMAN FILTER

A Kalman filter (KF) is one of the most frequently used algorithms to deal with estimating the unknown state over time. The estimation value continuously gets improved through observation [90]. KF is also known best for its ability to deal with sensor noise and applicable for any dynamic system which experiences consistent changes [22]. One of KF implementation benefits is that it does not require much data to be kept in the memory, and the only critical data is the previous state of the sensor signal. This issue proves that the KF suits real-time application systems with a minimum system specification requirement [91]. However, KF is limited to linear models of domain problems [92].

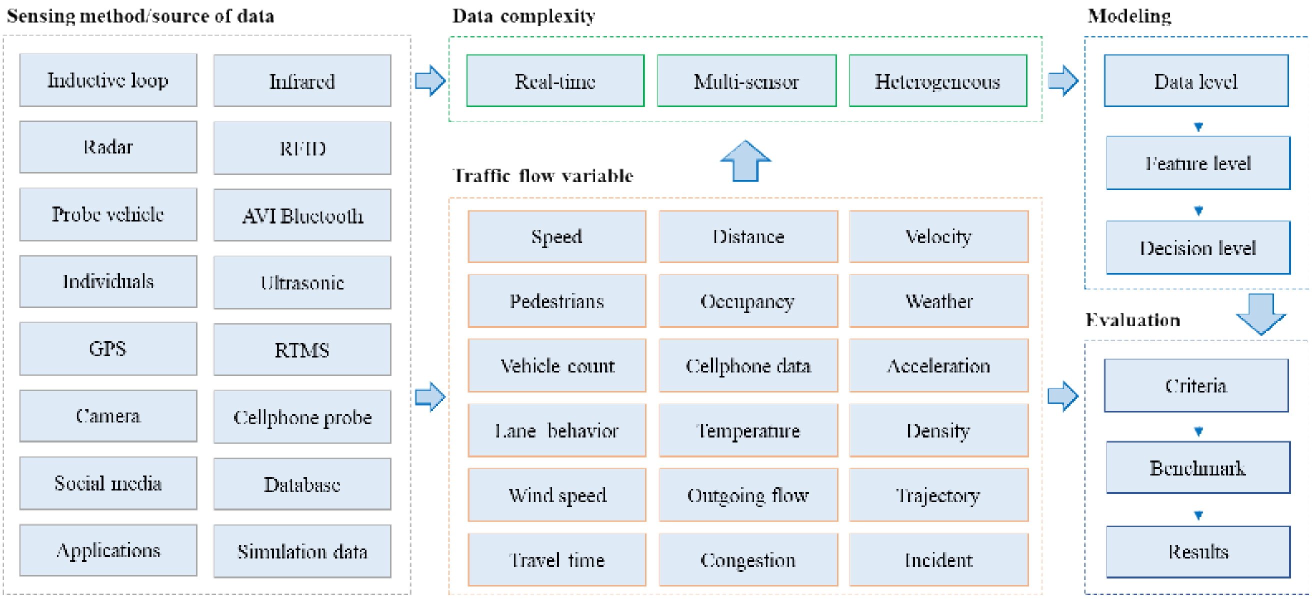



**FIGURE 3.** **Taxonomy of data fusion parameters in TFA systems.**

KF algorithm focuses on two processing stages, prediction and update. State estimation evolved from the previous updated state, and the updated state is determined by finding the difference between real measurement and estimation measurement [93]. Once updated state estimation is figured out, error covariance is calculated. A critical note here is that having a smaller value of error covariance than prediction error covariance indicates the state estimation parameter's measurement precision. This technique can also be used to perform estimation based on historical state and incorporate with current measurement. The state equation is shown below [12]:

$$x_t = A \cdot x_{t-1} + w_{t-1} \quad (1)$$

$$z_t = H \cdot x_t + v_t \quad (2)$$

$x_t$ indicate estimation of state, at time t,
$z_t$ indicate measurement at time t,
$w_t$ indicate noise process,
$v_t$ measurement noise
$A$ indicate state transfer matrix and
$H$ indicate parameter of the variable measured

This technique is suitable to make predictions based on currently collected heterogeneous data from multi-sensor [18]. Wang *et al.* [12] applied a KF technique to produce traffic state estimation based on the fusion process of GPS and RTMS data. The experiment is measured using accuracy to compare the accuracy during the off-peak and peak period. All the used four intersections show high accuracy during all conditions with Gaussian Mixture Model (GMM)–KF methods, off-peak (97%, 94%, 78%, and 94%), and peak period (94%, 92%, 91%, and 96%) compared to the other few tested methods.

Liu *et al.* [86] implement a progressive extended KF (PEKF) to produce traffic conditions estimation on heterogeneous data from multi-sensor. The sensors are microwave and wireless communication, which provide data of different characteristics. They introduce some inferences during the estimation process when there is a feature conflict from two different data sources. The absolute difference (AD) pattern shows that the estimation model is close to the real measurement.

Byon *et al.* [94] implement single-constraint-at-a-time (SCAAT) KF to fuse data from various sources and produce a better way of traffic conditions monitoring. The main concern here is to focus on a single recent measured sensor data to avoid any malfunction input device to interrupt the estimation process. They perform several test scenarios to evaluate the SCAAT KF. However, its performance efficiency is yet to be proven.

Zhang and Poschinger [95] apply an extended KF (EKF) based on the extended cell transmission model to estimate turning ratios using detector data. They implement two blocks of EKF-based filter that deal with real-time floating car and historical data. Nine experiments are conducted, and the variance of both filtered turning ratio one and filtered turning ratio two gives the best value based on the noise pair given during the third experiment. Cai *et al.* [39] use noise immune KF with non-Gaussian noise to perform traffic flow forecasting on real traffic data. Its immunity towards non-Gaussian noise is the main reason the technique is proposed. A dataset consists of four motorways named A1, A2, A4, and A8 are used in this study. The RMSE of the proposed model reduces to 38.83, 22.76, 24.65, and 24.46 percent. Ottaviano *et al.* [96] produce traffic estimation patterns with KF by integrating data from heterogeneous real-time data sources. In this study, multiple DF methods that are implemented have reduced the estimation error based on various incorporated data. After integrating GPS data into the KF model, the RMPSE is reduced to 7%. Saeedmanesh *et al.* [97] implement EKF to produce traffic estimation for a large-scale urban network by disaggregating a huge coverage area to multiple regions. Real-time measurement of each region is made by aggregating the accumulation before estimation is determined. They focus on producing state variables per region. Comparison between estimation and real demand (simulation) accuracies shows that the proposed model tends to follow a signal trend.

#### 2) NEURAL NETWORK

A neural network (NN) is a subset of algorithms that mimic the way the human brain operates [47]. NN comprises several different layers (input layer, hidden layers, and output layer) built up of nerve cells, and each node is paired to the next layer node [25]. It could work in an incomplete knowledge environment and perform parallel tasks simultaneously [98]. However, this technique requires accurate information and requires a device with higher processing power [47].

An *et al.* [40] conduct a prediction study by applying a fuzzy-based convolutional NN (F-CNN). F-CNN performs features extraction on the data, then training data to construct an improved CNN model and achieve the best results. The authors compare it with different methods such as SARIMA, VAR, DeepST, and RT-ResNet. Accuracy prediction of the proposed method is shown in RMSE performance by an improvement of 15.65%.

Peng *et al.* [41] propose a convolutional NN (CNN) with a deep learning model to generate traffic flow forecasting by using historical data of subway, taxi, and bus in Beijing. Evaluation and comparison among different methods that used the same dataset (TaxiBJ) are conducted. The proposed model manages to reduce RMSE from 0.0017 to 0.0008 and MAE from 0.00055 to 0.00031. Du *et al.* [99] proposed Fused Deep NN (F-DNN) to enhance pedestrian detection accuracy by introducing a pedestrian candidate generator. Evaluation conducted based on different settings of height and visibility of bounding box. In eight evaluation criteria, seven of them show ideal log-average miss rate (L-AMR) percentage level based on the proposed model.

Essien *et al.* [100] propose long short-term memory NN to improve traffic speed prediction by merging data from traffic and weather datasets. The model evaluation results have shown MAE value as 0.049, RMSE 0.0892, and MSE as 0.008. It shows that the environment factor plays a significant role in the traffic prediction model.

#### 3) DEMPSTER-SHAFER

Dempster-Shafer (DS) technique is known as a probability-based technique. It is used for classification based on mathematical theory, which is also known as the theory of belief functions [22]. The (0, 1) are the judgment values used to measure belief indication. DS is one of the suitable techniques when dealing with uncertainty and imprecise data [70]. DS is suitable for multi-sensor and heterogeneous data processing as it can evaluate the trust of different resources [101]. Each evidence is represented by a basic probability assignment which is also known as a density function. The relation between the degree of belief ($bel$) and plausibility ($pl$) is as below:

$$Pl\,(A) = 1 - bel(\neg A) \tag{3}$$

Wahab *et al.* [102] detect a misbehaved cluster of vehicles with the application of DS. Each node of the cluster helps in evidence aggregation before a decision can be made by the DS theory. DS plays an important role in filtering uncertain evidence and exclude them. The result shows improved detection ability for almost 40% and reduces selfish nodes up to 30%. Mehrannia *et al.* [49] make full use of evidence theory for fusing traffic sensor data received from various sensory systems available within a certain segment of the road and fuse the processed information to determine the area of road incident. The Best Performance in Scenario (BPS) is measured in this study, and the proposed technique is reliable to achieve better accuracy in different scenarios. The standard deviation is the variable for each scenario representation, such as normal traffic distribution, cars not well concentrated, and cars concentrated at the incident location.

Ning *et al.* [70] propose a reliability revaluation DS technique for real-time traffic estimation study. This technique could fuse heterogeneous traffic data of different formats like cameras and mobile phone data. The proposed method shows the lowest percentage of mean state decision error (MSDE), which is 6.9% for arteries and 8.3% for branches than three other methods. Soua *et al.* [88] propose DS as one of the fusing techniques to fuse decisions made for stream data that consists of traffic and weather data and event-based data that consists of traffic and tweet data. Few scenarios are tested to find a correlation between DS and deep belief network (DBN) as a feature level fusion technique to validate the DF model. The proposed model has shown 88.91% accuracy, 85.96% precision, 86.23% G-Mean, and 89.89% sensitivity. Fusion decision from the proposed framework has fulfilled a high level of accuracy compared to other classic methods. Gao *et al.* [45] use various data sources, including video data, GPS data, and logging device data, to trace lane-changing behavior among road drivers by assimilating improved DS based on correlation coefficient (DST-CC). The fused feature set from different modalities is combined in decision level fusion to determine lane change events. The model's evaluation results have shown an accuracy of 84.17% for 139 lane change left (LCL) event while 84.81% for 158 lane change right (LCR) event.

#### 4) FUZZY LOGIC

The idea behind Fuzzy Logic (FL) is to measure the degree of the state by stimulating human reasoning rather than judging the value based on absolute truth (1) or absolute false (0) scales [40], [103]. Some advantages of FL include its simplicity and ability to deal with imperfect data [104].

However, it requires enough understanding of possible scenarios that may arise [105]. The input signals generated by sensors can be in various forms. Each of the inputs or fuzzy sets is being mapped to certain real numbers. A membership value between [0,1] range is represented based on a mathematical basis to associate memberships to fuzzy variables in the fuzzification process. The fuzzification process is determined by rules that have been set and fuzzy inferences. Subsequently, the defuzzification is performed by functions such as the centroid that estimate a point over a fuzzy area to generate the final outputs.

Chen *et al.* [106] carry out a study on TFA prediction by applying fuzzy deep learning. This research integrates fuzzy representation in the deep learning model to reduce uncertain data. The proposed multilayer model obtained the best result with RMSE of 0.3037 and MRE of 0.2045 compared to few other popular algorithms such as ARIMA, DeepST, and CNN. Wang *et al.* [107] employ multilevel fuzzy theory to fuse features from real-time connected vehicle data to perform traffic condition evaluation. Within one hour of data, 92.6% of total packet data are classified as valid. Bouyahia *et al.* [108] applied fuzzy switching linear models to produce traffic state estimation from GPS data collected from the dataset. Evaluation of the proposed model based on open England traffic datasets shows 9% of the maximum absolute relative error.

#### 5) JOINT PROBABILISTIC DATA ASSOCIATION

Data association in DF implementation is a primary step in data preparation, and this is performed at the data level phase, where each incoming data must be received within time intervals. This task is performed before the estimation process takes place [22]. Joint probabilistic data association (JPDA), by the name itself, already indicates that data association is the primary goal. JPDA connects detected measurement within a specific interval of time with the target [109].

Various algorithms can be implemented under data association, for example, Multiple Target Tracking (MTT), Multiple Hypothesis Test (MHT), and Probabilistic Data Association (PDA) [22]. Joint association probabilities can be depicted as below [96]:

$$P(\theta \mid Z_k) = \frac{1}{K} p(z_k \mid \theta, X_k) P(\theta \mid X_k) \tag{4}$$

$K$ indicate normalization constant
$X_k$ indicate target state vector
$P(\theta \mid X_k)$ indicate the probability of assignment $\theta$

García *et al.* [110] compare JPDA over other data association techniques, which is the global nearest network. They incorporate two types of sensors and perform JPDA to the incoming sensor data to fuse data by applying the MTT approach to enhance the pedestrian detection mechanism. The proposed model shows 82.29% of positive detections with the lowest 1.11% misdetections per frame, compared to other devices and techniques discussed in this study. Hu *et al.* [111] used a joint probabilistic model to fuse movement data at lower-level detection to perform data classification on traffic scene prediction, engaging two hierarchical modules: upper and lower modules. Within ten frames of test scenario, RMSE shows 0.41, 0.33, 0.22, 0.14, 0.04 and 0.02 for each 4s, 3s, 2s, 1.5s, 1s and 0.5s interval.

#### 6) BAYESIAN

Bayesian is a statistical analysis technique that could produce the probability of an event or data source [37] by adopting the mathematical Bayes theorem of probability rule. This probability value is called posterior distribution, and this value summarizes the state of input by combining information from existing data with the help of a certain likelihood function. Bayesian is a simple yet powerful classification technique that works well for a large set of data. Bayes Theorem is shown below:

$$P(A \mid B) = \frac{\mathrm{P}(\mathrm{A})\, P(B \mid A)}{\mathrm{P}(\mathrm{B})} \tag{5}$$

$A, B$ indicate events
$P(A \mid B)$ indicate the probability of A given B is true
$P(B \mid A)$ indicate the probability of B given A is true
$P(A)$, $P(B)$ indicate independent probabilities of A and B

Zhang *et al.* [37] apply a combination of Bayesian methods to integrate multiple speed predictions from different traffic sources of data by fusing them to achieve speed estimation by considering a few traffic factors. The proposed model's evaluation shows better accuracy than other models: MAE 14.35, MAPE 5.92, and RMSE 9.77. Mil and Piantanakulchai [36] propose a framework of travel time estimation by combining Bayesian and GMM to enhance sensor accuracy data that contributes to travel time accuracy. From three case studies engaged in this study, MAPE shows the range of 3.46% to 16.3%.

Zhu *et al.* [112] propose a short-term traffic flow estimation model by applying Bayesian Network (BN), which takes spatial and temporal aspects into considerations. The prediction is performed based on variables and conditional distributions. The model shows a MAPE average of 18.2, 14.5, and 13.5 for five minutes, 10 minutes, and 15 minutes data from three different interval data. Liu *et al.* [34] carry out research on travel time estimation based on loop and probe vehicle data. They apply Bayesian fusion to fuse estimated travel time and develop an iterative estimation concept to enhance Bayesian estimation. The evaluation results show that the MAPE value is reduced to 4.8% and prove that Bayesian is a robust estimator.

#### 7) K-NEAREST NEIGHBOR

Clustering is a technique of grouping unlabeled data into a group based on certain characteristics of the data. The k-nearest neighbor (KNN) is one classification algorithm that can process features from numerical and categorical types [113]. The basic idea behind the KNN algorithm is to classify new data points received based on data characteristics defined from available data.

Distance between data (test data and training data) can be calculated in various ways, and one of them is using Euclidean distance. The formula is shown below [114]:

$$S\left(v_v, v_u\right) = \sqrt{\sum_{i=1}^{N_e}\sum_{j=1}^{N_o}\left(v_v\left(i,j\right) - v_u(i,j)\right)^2} \tag{6}$$

$S$ indicate the degree of similarity between two sets of data $v_v$ and $v_u$ indicate different sets of data

In recent studies, the KNN-based technique proves its ability to integrate with other DF techniques to improve the clustering process [115]. Yu *et al.* [114] develop a prediction model of short-term traffic conditions using the KNN algorithm. In this model, KNN groups each incoming data from a target road link, upstream and downstream, before the prediction process occurs. The proposed model shows the average of MAPEs of the first to fifth predictions as 10.37%, 13.58%, 16.81%, 19.49%, and 20.53%. Coluccia *et al.* [53] propose a set of detection procedures for radar detection by applying the KNN classifier algorithm. Tak *et al.* [35] make some improvements on KNN that is used for a traffic state prediction model. They introduce single-level and sequential search strategies to improve the data process and accuracy. Error distribution is measured and compared for the proposed model and traditional KNN algorithm. The proposed model shows that the error distribution is reduced (25% reduced to 2%) within the first 18 minutes compared to another algorithm that becomes stable after 27 minutes. Yu *et al.* [116] work on a study to generate short-term traffic state prediction. This model identifies the current pattern of traffic based on the historical state to make the prediction. Evaluation conducted shows that the proposed model achieves the best performance compared to another model with MAPE value roughly by 25.15%, 12.19%, 22.18%, 31.20%, 28.37%, 29.93%, and 27.62%.

#### 8) SOFTWARE AGENT

A software agent (SA) is a software component that can autonomously execute certain tasks, communicate, and share information within a system [117]. SA has intelligent capabilities to perform tasks based on certain plans and strategies. Its performance provides autonomous control to entities and systems. Each of these categories can instantiate different agents. An agent type depends on the field of study and the application domain that the agent is situated in [103]. The computational and task-specific are the most suitable agent types in DF studies. The flexibility brings great potential to achieve a modular-based system with less dependency and higher scalability [118]. However, the lack of efficient coordination between agents might affect overall system performance as this is one of the important characteristics of agent-based systems [119].

The modeling and implementation of an agent-based TFA system can be adapted according to the system setup and environment. Hamidi and Kamankesh [120] propose a three-layered agent-based system to perform traffic and transportation management during emergency conditions with consideration of traffic flow improvement. The result shows the positive impact on the road network's average speed when drivers use the proposed route. Tan *et al.* [121] highlight a local and supervisory agent concept in a hierarchical-based multi-agent system. A local agent has the capability to perform the best control based on the decision made within limited knowledge exploration, while a supervisory agent has a long-term learning experience to provide the best control decision to all local agents. The proposed model shows a 30.8% reduction of average delay compared to the classic model.

Xu *et al.* [122] develop multiple agents that are situated in a traffic control system to optimize traffic timing and enhance network performance of a road stretch for actuated control operation mode. This system is divided into two separate modules, and communication between them is in the form of a hierarchical multi-agent system (MAS). The proposed model shows a significant reduction of travel delay by 17.63%, 7.16%, and 8.25% over other methods while incrementing in average speed by 18.5%, 7.81%, and 8.62% over other methods.

#### 9) HYBRID ALGORITHM

Integrating data from various resources and large-scale systems require a robust model that can liaise with different data sources [69], [123]. Most AI-related technology and algorithms such as Deep Learning (DL), Machine learning (ML), and optimized tools can be utilized to achieve optimum results [124].

Few TFA related studies discuss AI contribution to DF framework when integrating with huge data from various data sources [9], [62], [63], [100], [88]. AI-related techniques have improved few areas of DF implementation in TFA related study, such as traffic flow prediction [41], [88], [106], [125], pedestrian detection module [126], traffic data classification [127], traffic speed prediction [38] and missing value estimation. Figure 4 shows a missing value estimation based on the DF technique in the training phase of the multimodal deep learning model (MDLM) of Li *et al.* [128].

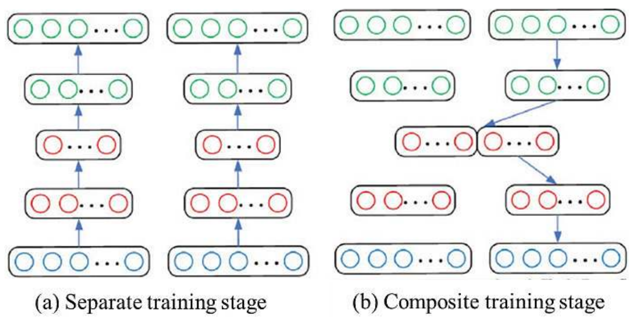



**FIGURE 4. DF Improvement training procedure of an MDLM [128].**

DL methods are known for the ability to manage the fusion process of heterogeneous data from multiple sources and robust enough to handle substantial unsupervised data from those multiple heterogeneous sources [69]. This technique can learn features with less preliminary knowledge and produce a certain pattern, which helps in decision-making [125].

On the other hand, ML provides a mechanism with which the machine or computer makes a certain discovery by learning from real-world data to gain an understanding of the pattern to perform a required task, and it is a continuous learning processing towards the huge data [127]. ML is proven to be one of the best classifier algorithms that link data from different sets of data by finding correlations among them and getting them converted into meaningful features [129]. Li *et al.* [38] implement ML by combining a few different DF techniques in the model to produce traffic speed prediction. Features are extracted from incoming data before the ML model performs its prediction. Four prediction models are compared, and the data level fusion shows that support vector regression performs well with a MAPE value of 7.85% and a RMSE of 8.71%. Figure 5 shows an example of a DF model based on a multi-agent system that controls several ML classifiers [119].

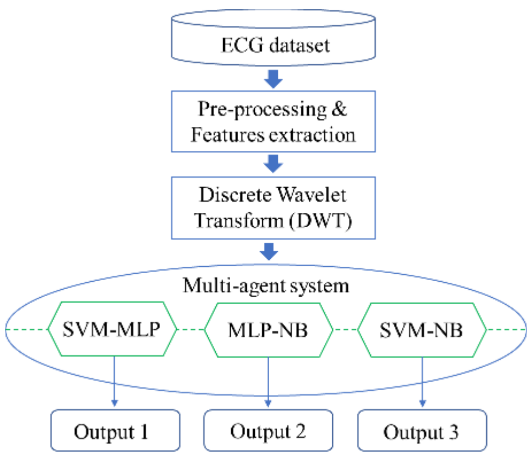



**FIGURE 5. Decision level fusion of ML classifiers [119].**

Koesdwiady *et al.* [125] implement a combination of DL, multi-task learning (MTL), and decision level by using connected vehicle data as well as weather information to get a better prediction of traffic flow. The proposed model improves accuracy by having an MAE of 0.0352, 0.0250, and 0.0654 for low traffic, medium traffic, and high traffic conditions, respectively, from the performance comparison matrix. The RMSE values for three conditions are 0.0481, 0.0356, and 0.0954. Li *et al.* [128] propose a multimodal DL model incorporating feature fusion to perform missing value estimation from heterogeneous traffic data. An experiment is conducted, and the correlation between missing data and accuracy is measured. The proposed model shows better accuracy by improving the RMSE by 49.38%, 32.20%, and 22.21% compared to the other three models.

Adetiloye and Awasthi [9] propose a big data fusion integration model to manage two different kinds of data from homogeneous and heterogeneous sources of data. This model consists of an ML algorithm to fuse quantitative data, such as backpropagation NN, RF, and DBN, to manage the same format of data. The heterogeneous model of fusion achieves a 0.1262 MAE value compared to other individual techniques. Gu *et al.* [130] propose the Bayesian technique's deployment to optimize the outputs of three types of DL prediction algorithms in the TFA model. The proposed model shows great accuracy from each aspect by achieving an MAE of 7.41, MAPE of 12.44%, and VAPE of 13.94% compared to the other seven models. The architecture of the improved DL-Bayesian encompasses a combination model that is used for short-term TFA prediction.

Nowak *et al.* [129] present a DF classifier module by combining Bayesian and decision tree to perform real-time processing data for a multi-sensor system environment. Based on the confusion matrix evaluation, the proposed model's error rate shows the lowest value, 3.41%, compared to the other two classifiers. Din *et al.* [131] propose a robust model of big data clustering with a routing technique to analyze data from multiple sensor systems. They divide the clustering algorithm into three phases: (1) setup phase, (2) steady phase, and (3) routing phase, and they prove efficient results at the end of the study. Table 4 shows some examples of hybrid models in the TFA study.

**TABLE 4. Examples of hybrid models in TFA.**

| Methods | Result to achieve | Ref. |
|---|---|---|
| Backpropagation NN, RF, DBN | Big DF for traffic congestion | [9] |
| EKF, Sentiment analysis, cluster | Traffic congestion prediction | [9] |
| FL, CNN | Traffic flow prediction | [40] |
| DL, NN | Big DF for traffic speed | [100] |
| DL, DS | Traffic flow prediction | [88] |
| DL, FL | Traffic flow prediction | [106] |
| RF, ML | Traffic accident detection | [132] |
| Bayesian, Decision tree | Data classification of real-time data | [129] |
| Bayesian, DL | Traffic flow prediction | [130] |

Performing prediction or estimation based on a single method has its own limitation when aggregating linear and non-linear models [42]. Higher accuracy results can be achieved by various sources integration to the combination algorithm model [37].

## B. EVALUATION TECHNIQUES AND CRITERIA

Research outcome requires certain measurement tools to define the effectiveness of DF models, either quantitatively or qualitatively. Shen and Zhu [133] highlight the need for evaluation after performing the estimation fusion to measure the performance efficiency level. Linear or non-linear process of selected DF method depends on the data properties, proposed model, and the domain issue being modeled. Four techniques that are briefed in this section are root mean square error (RMSE), root mean percentage square error (RMPSE), mean absolute error (MAE) and mean absolute percentage error (MAPE).

Cai *et al.* [134] apply MAPE and RMSE to measure the short-term traffic forecasting model to find the relative errors. Cai *et al.* [39] also evaluate the proposed forecasting model with MAPE and RMSE using benchmark datasets. Kawasaki *et al.* [65] choose MAPE to measure the average error ratio in the estimation result. Mil and Piantanakulchai [72] estimate their proposed model's travel time error by adopting MAPE and MAE.

RMSE is more reactive towards outliers and suitable for a normal error distribution [135]. Since most of the prediction model tends to produce normal distribution, RMSE would be the most suitable choice [136]. On the other hand, MAPE is used when there are no extreme or zero values found in the data [137], [138]. Jiang *et al.* [17] apply RMSE to evaluate a proposed traffic state estimation model's performance. Seo and Kusakabe [73] choose MAPE and RMSE to weigh the error in traffic density estimation. Li *et al.* [38] use MAPE and RMSE to compare traffic speed prediction value's effectiveness with ground truth values. Deng *et al.* [54] measure their traffic estimation effectiveness value by using the MAPE evaluation metric.

#### 1) ROOT MEAN SQUARE ERROR (RMSE)

Comparing bias of prediction over the actual result is a frequently used technique to evaluate the proposed models' accuracy [139], [140]. It can be measured over time based on the type and source of the implemented data [134]. This introduces the concept of root mean square error (RMSE), which is a quite common and standardized measure for regression analysis. It is the standard deviation of the prediction errors by determining the estimation error variance [141]. RMSE provides a clear picture of the error scattering [136]. The formula to measure RMSE is as below [68]:

$$RMSE = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(x_n - \hat{}x_n)^2} \tag{7}$$

#### 2) ROOT MEAN PERCENTAGE SQUARE ERROR (RMPSE)

Root Mean Percentage Square Error (RMPSE) calculation has the same characters as RMSE, or known as an extension of RMSE, by taking the mean of the target result to form a percentage value. This value indicates the magnitude percentage error of the value from the result calculated. The formula to measure RMPSE as below [68]:

$$RMPSE = 100 \times \sqrt{\frac{1}{N}\sum_{n=1}^{N}\frac{(x_n - \hat{}x_n)^2}{x_n^2}} \tag{8}$$

#### 3) MEAN ABSOLUTE ERROR (MAE)

Other than RMSE, MAE is another most common way to measure variables' accuracy. It is used to calculate the average errors in a set of estimated values over actual values. This evaluation technique is scale-independent [140] and extremely useful for model evaluations and gives the same weight to the error [136]. The formula to calculate MAE as below [128]:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|e_i| \tag{9}$$

#### 4) MEAN ABSOLUTE PERCENTAGE ERROR (MAPE)

Mean Absolute Percentage Error (MAPE) is another way of measuring the effectiveness of research result implementation and is known as the standard of indicating errors [139], [140]. This measurement tool is used to measure forecasted accuracy by calculating the average of normalized absolute error [54] by finding the average ratio over the data. This value represents the mean absolute percentage error value between actual and predicted value in percentage form [140], [141]. The formula to calculate MAPE as below [68]:

$$MAPE = 100 \times \frac{1}{N}\sum_{n=1}^{N}\frac{|x_n - \hat{}x_n|}{x_n} \tag{10}$$

## V. ANALYSIS AND DISCUSSION

DF in TFA related studies has contributed significant improvement to the ITS field such as travel time estimation [72], [74], [77], traffic incident detection [49], [142], [143], driver's lane-changing behavior [45], traffic state estimation [65], [70], [71] road/highway management [13], [52], [54], smart traffic light system [144], optimizing pedestrian detection [110], [99] and traffic emergency management [145]. System variables like state estimation for freeways, highways, and speed are valuable for traffic light operation improvement, surveillance, management and to improve human decision-making [21], [146]. Forecasting and traffic incident are other outputs of DF in TFA related applications, while incident detection could help in emergency response improvement [1], [21], [147]. Detection procedures for vehicles and pedestrians can improve DF methods manipulation to achieve better detection accuracy [53], [99], [110], [148].

### A. ANALYSIS

A few DF methods or techniques implemented in TFA related studies are discussed in Section IV. Table 5 shows these methods, the relationships between these methods, the related abstraction level of fusion, the testing datasets and their characteristics, the evaluation variables and criteria, and the outcome to be achieved.

Based on Table 5, we observe that the KF is mainly applied for performing feature level fusion. It is used in traffic state estimation [12], [86] and vehicle turning ratio estimation [95]. NN is best served as a feature level fusion technique to produce flow prediction [40], [41], and filtering pedestrian detection [99]. DS works best as a classifier at decision level fusion to detect lane change activity [45], detect misbehaved clusters of vehicles [102], and traffic state estimation [70]. FL may serve either one or both feature level and decision level fusion to achieve traffic prediction [106] and estima-

**TABLE 5.** DF methods in TFA.

| DF Method | Abstraction Level Fusion | Dataset | | Evaluation | | Outcome | Ref. |
|---|---|---|---|---|---|---|---|
| | | Testing data | Characteristic | Variable | Criteria | | |
| KF | Feature level | Four segments road in Hangzhou | Multi-sensor, heterogeneous | Average speed, average occupancy (congestion level) | Accuracy | Traffic state estimation | [12] |
| | Feature level | Jiangsu freeway province, China | Multi-sensor, heterogeneous, real-time | Traffic flow, speed, cellphone counts | Absolute difference | Traffic condition estimation | [86] |
| | Feature level | VISSIM traffic simulation | Heterogeneous | Traffic flow, density | Variance | Vehicle turning ratio estimation | [95] |
| NN | Feature level | Beijing taxicab (TaxiBJ) trajectory and meteorology data | Multi-sensor, heterogeneous, real-time | GPS, temperature, wind speed, date | RMSE, MSE, MAE | Traffic flow prediction | [40] |
| | Feature level | Beijing subway (SubwayBJ) and Beijing bus (BusBJ) | Multi-sensor, heterogeneous, real-time | Check-in & check-out time, getting-in & getting of time, GPS | RMSE, MAE | Traffic passenger flow prediction | [41] |
| | Feature level | Caltech Pedestrian dataset | Heterogeneous | Bounding box, frame | L-AMR | To filter pedestrian candidate | [99] |
| DS | Decision level | Detector and video cameras from Guangzhou Traffic Police Headquarters, and GPS data from Guangzhou Communications Commission of Municipality | Multi-sensor, heterogeneous, real-time | GPS, speed | MSDE | Traffic state estimation | [70] |
| | Decision level | MATLAB network simulator and VanetMobiSim traffic simulator | Multi-sensor, heterogeneous, real-time | Speed, velocity distance, GPS | Probability of detection | Determine misbehave cluster of vehicles | [102] |
| | Decision level | Ten driver's real driving around University Michigan, Arbor and Dearborn campus | Multi-sensor, heterogeneous, real-time | GPS, wheel odometry, onboard diagnostic | Confidence, accuracy | Detect lane change event | [45] |
| FL | Feature level | Beijing Taxi (TaxiBJ) and TaxiCab GPS | Multi-sensor, heterogeneous, real-time | GPS, input flow, output flow, weather, temperature | MAE, RMSE, MRE | Traffic flow prediction | [106] |
| | Feature level | Real-time data of experimental vehicles, Pingguoyuan South Road, Shijingshan District of Beijing | Heterogeneous, real-time | GPS, number of lanes, road length | Multilevel fuzzy synthetic | Traffic condition estimation | [107] |
| | Decision level | Online highway data of England expressways (Jan 1st, 2014 – Jan 31st, 2014) | Multi-sensor, heterogeneous, real-time | Average density, speed, GPS | MRE | Traffic state estimation | [108] |
| JPDA | Data level | Real-world devices (a laser scanner, a camera, and a car) | Heterogeneous, real-time | GPS, pedestrian detection | Positive detection rate, misdetection rate | Fuse detection data | [110] |
| | Data level | Computer simulations | Multi-sensor, heterogeneous | Velocity, detection | MSE | Fuse and associate data | [149] |
| | Data level | NGSIM | Multi-sensor, heterogeneous, real-time | Trajectory, velocity, acceleration | RMSE | Fuse motion models | [111] |
| Bayesian | Feature level | VISSIM simulation on 18 links and eight intersections | Multi-sensor, heterogeneous, real-time | GPS, volume, speed, occupancy | MAPE | To estimate travel time | [34] |
| | Feature level | California I-880 corridor | Multi-sensor, heterogeneous, real-time | GPS, speed | MAPE | To estimate travel time | [36] |
| | Feature level | Microscopic traffic simulation dataset | Heterogeneous, real-time | Detection, volume, speed | MAPE | To estimate short-term traffic flow | [112] |
| | Feature level | Interstate 5 (I-5) freeway in the northwest region of Washington State | Multi-sensor, heterogeneous, real-time | Volume, speed, average vehicle length, timestamp | MAE, MAPE, RMSE | To estimate speed by integrating data from different sources | [37] |
| KNN | Data level | Kyungbu expressway, Republic of Korea | Multi-sensor, heterogeneous, real-time | GPS, speed, vehicle volume | Error distribution | Travel time prediction | [35] |
| | Data level | GPS data of taxis in Foshan, China. 6 road links on Foshan Avenue | Multi-sensor, heterogeneous, real-time | Delay, GPS, speed, velocity | MAPE | Fuse road data upstream and downstream direction | [114] |
| | Data level | Beijing Wroker's Stadium | Heterogeneous, real-time | Average speed, GPS | MAE, MAPE | Traffic state prediction | [116] |
| SA | Decision level | Simulation (18 one-way street and 43 two-way street) | Heterogeneous, real-time | Vehicle volume, traffic flow, speed | Average speed of vehicles | Traffic management transportation | [120] |
| | Decision level | Simulation of an arterial traffic network | Heterogeneous, real-time | Traffic state, cycle time fraction | Average delay reduction | Traffic control decision | [121] |
| | Decision level | Simulation (MATLAB and VISSIM) | Heterogeneous, real-time | Traffic flow, speed, queue length | Travel delay reduction | Traffic control signal | [122] |
| Hybrid Algorithm | Feature and decision levels | Genetec blufaxcloud travel-time system engine and Twitter | Multi-sensor, heterogeneous, real-time | Travel time, | MAE, R | Traffic congestion prediction | [9] |
| | Feature level | Transport for Greater Manchester (TfGM) | Multi-sensor, heterogeneous, real-time | Average speed, flow, density | MAE, RMSE, MSE | Traffic speed prediction | [100] |
| | Feature level | Caltrans Performance Measurement Systems (PeMS), MesoWest project, and CityPulse Dataset Collection | Multi-sensor, heterogeneous, real-time | Traffic flow, weather, event (social media data) | RMSE, MAE | Traffic flow prediction | [88] |
| | Feature level | Beijing Taxi (TaxiBJ) | Multi-sensor, heterogeneous, real-time | GPS, input flow, output flow | MAE, MRE, RMSE | Traffic flow prediction | [106] |

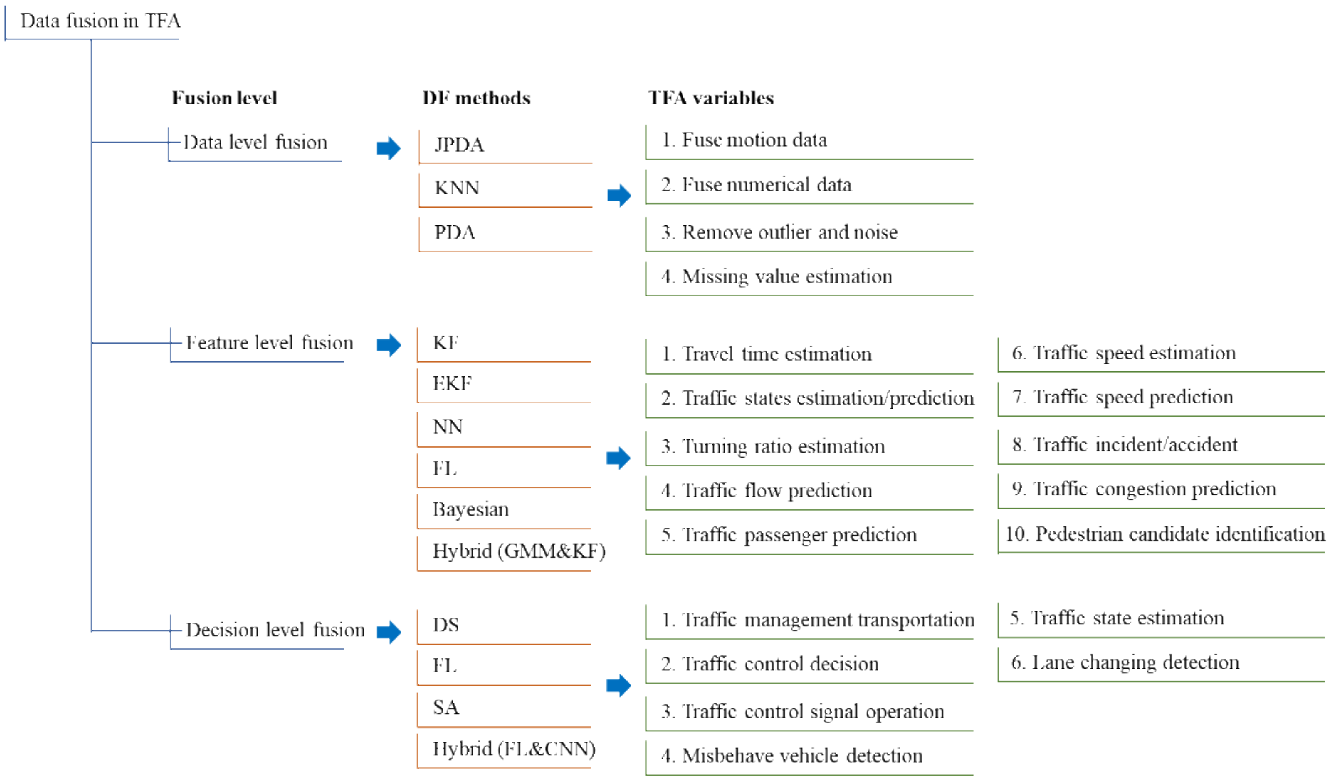



**FIGURE 6.** Taxonomy of data fusion methods in TFA systems.

tion [107], [108]. JPDA is mainly implemented at data level fusion and used mostly during the early stage of data processing in traffic data association [110], [149], and motions [111]. Bayesian works well in achieving estimation value in travel time [34], [36], speed [37], and traffic flow [112]. KNN is another method that performs well at data level fusion to produce travel time [35], traffic state prediction [116], and fusing incoming traffic direction [114]. SA proves its ability at decision level fusion in traffic light management system [120]–[122]. The flexibility of the hybrid models that they work at feature level and decision level fusions. They are implemented to achieve traffic congestion prediction [9], traffic speed prediction [100], and traffic flow prediction [88], [106]. Figure 6 summarizes the DF taxonomy in TFA studies.

Integration among the various sources of data ensures higher accuracy and optimized results [37]. This is clearly presented in the review of Table 5. All the TFA study focuses on one or more combinations of multi-sensor, heterogeneous and real-time data. A few studies conducted focus on improving highway efficiency and dynamic estimation by incorporating data from different sensors [14], [52], [96]. Adetiloye and Awasthi [9] propose a traffic congestion prediction model by integrating two different kinds of data, homogeneous and heterogeneous, from various resources. A specific DF technique handles each 'block' of data before features are used to achieve the result. An *et al.* [40] make full use of the weather, wind speed, and GPS data to perform traffic flow prediction. Akbar *et al.* [64] proposed a model for congestion prediction by considering traffic data, Twitter data, and weather data as input streams. Extensive exploration of hybrid models brings a positive impact to the TFA study. They provide more dynamic models to deal with different kinds of system environments, including prediction [146], [147], fixing missing values from heterogeneous data [128], and post-impact of an incident [148]. SA improves DF model performance at a higher level and considers all required data and specifications [120].

### B. OPEN RESEARCH DIRECTIONS

The ability to gather data from multiple sources or/and multimodal traffic sensing devices improves numerous areas in ITS development. However, the challenges or complexities in the model implementation are undeniable. Figure 7 summarizes the identified challenges of DF-TFA research and its applications from this review. The review yields five challenging categories and their related issues. The categories are devices, data preprocessing, research, system architecture, and processing complexity, and their description is as below:

#### 1) DEVICE ISSUES

Adopting multi-device requires additional indexing, synchronization, and mapping skills and methods [80] to ensure the device's reliability as data collection tools. However, there is a cost involved to ensure device stability and maintenance [2].

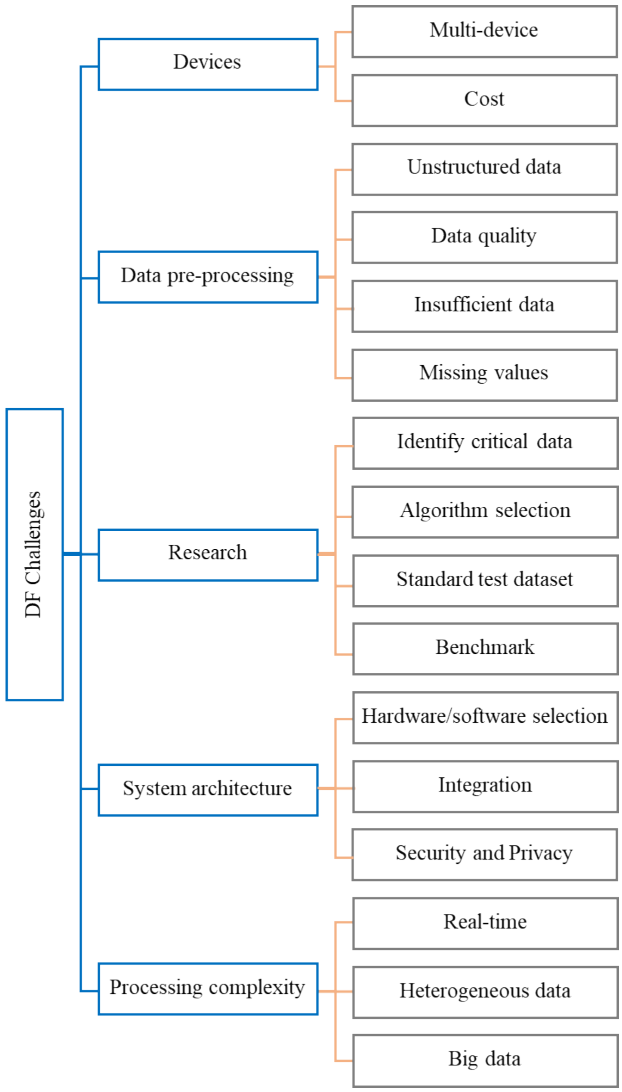



**FIGURE 7.** DF challenges in TFA domains.

### 2) DATA PRE-PROCESSING ISSUES

Huge and unstructured data is hard to manage and requires complex preprocessing methods [150], [151]. Dealing with those data in a multi-sensor and heterogeneous sources environment is the real challenge [60], [69]. Data quality and missing values from the datasets need to be taken care of to avoid inaccurate results [64], [152].

### 3) RESEARCH ISSUES

Identifying critical data to fit the research purpose is a real challenge [60], [150]. The diversity of algorithms and methods require intensive study to formulate a suitable model to solve domain problems [9], [61]. Finding the complete, suitable, and relevant dataset is another challenge when evaluating DF proposed models or solutions, especially for a data-dependent model [25]. Performing prediction or estimation kind of output in TFA study requires benchmarking to evaluate and validate the performance and ability of the DF model [130]. In DF related studies, it is difficult to have a benchmark at the model level. Still, it is possible to compare similar purpose algorithms such as ML algorithms for classification or prediction tasks.

### 4) SYSTEM ARCHITECTURE ISSUES

Hardware specification requires compatible software to be integrated to ensure system performance and stability [153]. Integration of different types of devices, sensors, algorithms, and methods has its own complexities. Collecting data from various sources and inputting them into the DF model requires integration with several systems that are equipped with stable communication networks [154]. In cloud computing as a data center, security and privacy leakage are other key areas that need to be focused on [19], [27].

### 5) PROCESSING COMPLEXITY ISSUES

Collecting data from various sources and processing them in a real-time environment requires complex distributed and dynamic systems [154]. Heterogeneous data sources increase data completeness and reliability. However, dealing with heterogeneous data with different characteristics may require a model combination as a solution that possibly increases the complexity [69], [61].

There are various ways of fixing the current challenges of DF implementation. Multiple DF method combination or hybrid algorithm is one way of dealing with complex and unstructured data [61]. Flexible system architecture with suitable technology can provide the means to overcome the challenges of modeling DF for TFA, like reducing the complexity of such systems. The source of data is not limited only to traffic sensors as discussed in the previous section, but recently, researchers also exploit more unstructured and dynamic data from social media networks, such as Twitter [9], [64], [147], and Instagram [155].

Alkouz and Aghbari [155], in their recent study, develop a traffic jam prediction classifier by merging data from different sources of social media feeds in different lingual. This indicates, more dynamic data from various data sources has a bright potential to be incorporated in DF of TFA related studies in the future. Recently, Li *et al.* [156] conduct specific research about trajectory data in the latest development for traffic study purposes. IoT helps in integrating data from different methods for various purposes [27]. Akbar *et al.* [64] work on a study to analyze IoT data by integrating them (traffic, weather, and Twitter data) to predict real-time congestion. Mei *et al.* [157] make full use of crowd-sensing traffic vehicle data to provide a city monitoring mechanism. Mai-Tan *et al.* [158] propose an architecture of crowdsourcing data for traffic estimation, which consists of various data from monitoring systems, public websites, and mobile data collection.

## VI. CONCLUSION

This paper attempts to provide an insight into the proposed DF model specifically in TFA related studies. Production of this review paper entails carefully analyzing the content of 158 articles from the year 2010 until 2020. The review includes 139 research articles and 19 review articles. This review offers a comprehensive insight into DF modeling and implementation in TFA. It describes the evaluation methods and criteria being applied according to the incorporated traffic parameters. Moreover, this paper attempts to link each fusion level with some suitable DF methods and traffic variables to provide a clear insight into the way DF is being utilized in various ITS studies. Each DF method has its specific role either at data level, feature level, or decision level fusion. The data sources that a DF model deals with specify the complexity and challenges associated with it. The complexity results from multi-sensor, heterogeneous and real-time system environments. Continuous exploration of hybrid solution models of DF might lead to a bright future in various TFA fields. Subsequently, intensive reviews should be initiated to focus on DF's involvement in these areas in order to innovate better solutions.

## REFERENCES

[1] N.-E.-E. Faouzi, H. Leung, and A. Kurian, "Data fusion in intelligent transportation systems: Progress and challenges—A survey," *Inf. Fusion*, vol. 12, no. 1, pp. 4–10, Jan. 2011.

[2] J. Guerrero-Ibáñez, S. Zeadally, and J. Contreras-Castillo, "Sensor technologies for intelligent transportation systems," *Sensors*, vol. 18, no. 4, p. 1212, Apr. 2018.

[3] C. Bachmann, B. Abdulhai, M. J. Roorda, and B. Moshiri, "A comparative assessment of multi-sensor data fusion techniques for freeway traffic speed estimation using microsimulation modeling," *Transp. Res. C, Emerg. Technol.*, vol. 26, pp. 33–48, Jan. 2013.

[4] A. Vaccaro, P. Mercogliano, P. Schiano, and D. Villacci, "An adaptive framework based on multi-model data fusion for one-day-ahead wind power forecasting," *Electr. Power Syst. Res.*, vol. 81, no. 3, pp. 775–782, Mar. 2011.

[5] G. Fortino, S. Galzarano, R. Gravina, and W. Li, "A framework for collaborative computing and multi-sensor data fusion in body sensor networks," *Inf. Fusion*, vol. 22, pp. 50–70, Mar. 2015.

[6] T. Kusakabe and Y. Asakura, "Behavioural data mining of transit smart card data: A data fusion approach," *Transp. Res. C, Emerg. Technol.*, vol. 46, pp. 179–191, Sep. 2014.

[7] A. Ortiz, D. Rosario, O. Fuentes, and S. Blair, "Image-based 3D model and hyperspectral data fusion for improved scene understanding," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2017, pp. 4020–4023.

[8] J. Kim, L. Kim, and H. Sohn, "Autonomous dynamic displacement estimation from data fusion of acceleration and intermittent displacement measurements," *Mech. Syst. Signal Process.*, vol. 42, nos. 1–2, pp. 194–205, 2014.

[9] T. Adetiloye and A. Awasthi, "Multimodal big data fusion for traffic congestion prediction," in *Multimodal Analytics for Next-Generation Big Data Technologies and Applications*. Cham, Switzerland: Springer, 2019, pp. 319–335.

[10] X. Chen, S. Zhang, L. Li, and L. Li, "Adaptive rolling smoothing with heterogeneous data for traffic state estimation and prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 2, pp. 1247–1258, Apr. 2019.

[11] M. Rahmani, E. Jenelius, and H. N. Koutsopoulos, "Floating car and camera data fusion for non-parametric route travel time estimation," in *Proc. 17th Int. IEEE Conf. Intell. Transp. Syst. (ITSC)*, Oct. 2014, pp. 1286–1291.

[12] C. Wang, Q. Zhu, Z. Shan, Y. Xia, and Y. Liu, "Fusing heterogeneous traffic data by Kalman filters and Gaussian mixture models," in *Proc. 17th Int. IEEE Conf. Intell. Transp. Syst. (ITSC)*, Oct. 2014, pp. 276–281.

[13] Z. Shan, Y. Xia, P. Hou, and J. He, "Fusing incomplete multisensor heterogeneous data to estimate urban traffic," *IEEE Multimedia*, vol. 23, no. 3, pp. 56–63, Jul./Sep. 2016.

[14] N. Bekiaris-Liberis, C. Roncoli, and M. Papageorgiou, "Highway traffic state estimation with mixed connected and conventional vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 12, pp. 3484–3497, Dec. 2016.

[15] A. Nantes, D. Ngoduy, A. Bhaskar, M. Miska, and E. Chung, "Real-time traffic state estimation in urban corridors from heterogeneous data," *Transp. Res. C, Emerg. Technol.*, vol. 66, pp. 99–118, May 2016.

[16] T. Seo, T. Kusakabe, and Y. Asakura, "Traffic state estimation with the advanced probe vehicles using data assimilation," in *Proc. IEEE 18th Int. Conf. Intell. Transp. Syst.*, Sep. 2015, pp. 824–830.

[17] Z. Jiang, X. Chen, and Y. Ouyang, "Traffic state and emission estimation for urban expressways based on heterogeneous data," *Transp. Res. D, Transp. Environ.*, vol. 53, pp. 440–453, Jun. 2017.

[18] F. Alam, R. Mehmood, I. Katib, N. N. Albogami, and A. Albeshri, "Data fusion and IoT for smart ubiquitous environments: A survey," *IEEE Access*, vol. 5, pp. 9533–9554, 2017.

[19] J. A. Guerrero-ibanez, S. Zeadally, and J. Contreras-Castillo, "Integration challenges of intelligent transportation systems with connected vehicle, cloud computing, and Internet of Things technologies," *IEEE Wireless Commun.*, vol. 22, no. 6, pp. 122–128, Dec. 2015.

[20] P. H. L. Rettore, R. Rigolin F. Lopes, G. Maia, L. A. Villas, and A. A. F. Loureiro, "Towards a traffic data enrichment sensor based on heterogeneous data fusion for ITS," in *Proc. 15th Int. Conf. Distrib. Comput. Sensor Syst. (DCOSS)*, May 2019, pp. 570–577.

[21] N.-E. El Faouzi and L. A. Klein, "Data fusion for ITS: Techniques and research needs," *Transp. Res. Procedia*, vol. 15, pp. 495–512, Jan. 2016.

[22] F. Castanedo, "A review of data fusion techniques," *Sci. World J.*, vol. 2013, Sep. 2013, Art. no. 704504.

[23] R. Wang, W. Ji, M. Liu, X. Wang, J. Weng, S. Deng, S. Gao, and C.-A. Yuan, "Review on mining data from multiple data sources," *Pattern Recognit. Lett.*, vol. 109, pp. 120–128, Jul. 2018.

[24] E. Azimirad, J. Haddadnia, and A. Izadipour, "A comprehensive review of the multi-sensor data fusion architectures," *J. Theor. Appl. Inf. Technol.*, vol. 71, pp. 33–41, Jan. 2015.

[25] T. Meng, X. Jing, Z. Yan, and W. Pedrycz, "A survey on machine learning for data fusion," *Inf. Fusion*, vol. 57, pp. 115–129, May 2020.

[26] M Jevtić, N. Zogović, and S. Graovac, "Multi-sensor data fusion architectures revisited," in *Proc. ICIST*, vol. 1, 2019, pp. 119–123.

[27] W. Ding, X. Jing, Z. Yan, and L. T. Yang, "A survey on data fusion in Internet of Things: Towards secure and privacy-preserving fusion," *Inf. Fusion*, vol. 51, pp. 129–144, Nov. 2019.

[28] J. Liu, T. Li, P. Xie, S. Du, F. Teng, and X. Yang, "Urban big data fusion based on deep learning: An overview," *Inf. Fusion*, vol. 53, pp. 123–133, Jan. 2020.

[29] S. Jusoh and S. Almajali, "A systematic review on fusion techniques and approaches used in applications," *IEEE Access*, vol. 8, pp. 14424–14439, 2020.

[30] S. A. Mostafa, A. Mustapha, M. A. Mohammed, M. S. Ahmad, and M. A. Mahmoud, "A fuzzy logic control in adjustable autonomy of a multi-agent system for an automated elderly movement monitoring application," *Int. J. Med. Informat.*, vol. 112, pp. 173–184, Apr. 2018.

[31] Z. Li, Y. Zhu, H. Zhu, and M. Li, "Compressive sensing approach to urban traffic sensing," in *Proc. 31st Int. Conf. Distrib. Comput. Syst.*, Jun. 2011, pp. 889–898.

[32] T. Seo, T. Kusakabe, and Y. Asakura, "Estimation of flow and density using probe vehicles with spacing measurement equipment," *Transp. Res. C, Emerg. Technol.*, vol. 53, pp. 134–150, Apr. 2015.

[33] A. E. Papacharalampous, O. Cats, J.-W. Lankhaar, W. Daamen, and H. Van Lint, "Multimodal data fusion for big events," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2549, pp. 118–126, Dec. 2016.

[34] K. Liu, M.-Y. Cui, P. Cao, and J.-B. Wang, "Iterative Bayesian estimation of travel times on urban arterials: Fusing loop detector and probe vehicle data," *PLoS ONE*, vol. 11, no. 6, Jun. 2016, Art. no. e0158123.

[35] S. Tak, S. Kim, K. Jang, and H. Yeo, "Real-time travel time prediction using multi-level $k$-nearest neighbor algorithm and data fusion method," in *Proc. Comput. Civil Building Eng.*, 2014, pp. 1861–1868.

[36] S. Mil and M. Piantanakulchai, "Modified Bayesian data fusion model for travel time estimation considering spurious data and traffic conditions," *Appl. Soft Comput.*, vol. 72, pp. 65–78, Nov. 2018.

[37] W. Zhang, Y. Qi, Z. Zhou, S. A. Biancardo, and Y. Wang, "Real-time travel time prediction using multi-level $k$-nearest neighbor algorithm and data fusion method," *IET Intell. Transp. Syst.*, vol. 12, no. 10, pp. 1312–1321, Dec. 2018.

[38] L. Li, X. Qu, J. Zhang, Y. Wang, and B. Ran, "Traffic speed prediction for intelligent transportation system based on a deep feature fusion model," *J. Intell. Transp. Syst.*, vol. 23, no. 6, pp. 605–616, 2019.

[39] L. Cai, Z. Zhang, J. Yang, Y. Yu, T. Zhou, and J. Qin, "A noise-immune Kalman filter for short-term traffic flow forecasting," *Phys. A, Stat. Mech. Appl.*, vol. 536, Dec. 2019, Art. no. 122601.

[40] J. An, L. Fu, M. Hu, W. Chen, and J. Zhan, "A novel fuzzy-based convolutional neural network method to traffic flow prediction with uncertain traffic accident information," *IEEE Access*, vol. 7, pp. 20708–20722, 2019.

[41] H. Peng, H. Wang, B. Du, M. Z. A. Bhuiyan, H. Ma, J. Liu, L. Wang, Z. Yang, L. Du, S. Wang, and P. S. Yu, "Spatial temporal incidence dynamic graph neural networks for traffic flow forecasting," *Inf. Sci.*, vol. 521, pp. 277–290, Jun. 2020.

[42] B. Alsolami, R. Mehmood, and A. Albeshri, "Hybrid statistical and machine learning methods for road traffic prediction: A review and tutorial," in *Smart Infrastructure and Applications*. Cham, Switzerland: Springer, 2020, pp. 115–133.

[43] A. Zeroual, F. Harrou, and Y. Sun, "Road traffic density estimation and congestion detection with a hybrid observer-based strategy," *Sustain. Cities Soc.*, vol. 46, Apr. 2019, Art. no. 101411.

[44] A. Bhaskar, T. Tsubota, L. M. Kieu, and E. Chung, "Urban traffic state estimation: Fusing point and zone based data," *Transp. Res. C, Emerg. Technol.*, vol. 48, pp. 120–142, Nov. 2014.

[45] J. Gao, Y. L. Murphey, and H. Zhu, "Personalized detection of lane changing behavior using multisensor data fusion," *Computing*, vol. 101, no. 12, pp. 1837–1860, Dec. 2019.

[46] A. Chatterjee, A. M. Michalak, R. A. Kahn, S. R. Paradise, A. J. Braverman, and C. E. Miller, "A geostatistical data fusion technique for merging remote sensing and ground-based observations of aerosol optical thickness," *J. Geophys. Res.*, vol. 115, no. D20, pp. 1–12, 2010.

[47] S.-F. Jiang, C.-M. Zhang, and S. Zhang, "Two-stage structural damage detection using fuzzy neural networks and data fusion techniques," *Expert Syst. Appl.*, vol. 38, no. 1, pp. 511–519, Jan. 2011.

[48] B. R. Chang, H. F. Tsai, and C.-P. Young, "Intelligent data fusion system for predicting vehicle collision warning using vision/GPS sensing," *Expert Syst. Appl.*, vol. 37, no. 3, pp. 2439–2450, 2010.

[49] P. Mehrannia, A. A. Moghadam, and O. A. Basir, "A dempster-shafer sensor fusion approach for traffic incident detection and localization," in *Proc. 21st Int. Conf. Intell. Transp. Syst. (ITSC)*, Nov. 2018, pp. 3911–3916.

[50] M. Aqib, R. Mehmood, A. Alzahrani, and I. Katib, "In-memory deep learning computations on GPUs for prediction of road traffic incidents using big data fusion," in *Smart Infrastructure and Applications*. Cham, Switzerland: Springer, 2020, pp. 79–114.

[51] S. B. Ayed, H. Trichili, and A. M. Alimi, "Data fusion architectures: A survey and comparison," in *Proc. 15th Int. Conf. Intell. Syst. Design Appl. (ISDA)*, Dec. 2015, pp. 277–282.

[52] T. Seo, A. M. Bayen, T. Kusakabe, and Y. Asakura, "Traffic state estimation on highway: A comprehensive survey," *Annu. Rev. Control*, vol. 43, pp. 128–151, Jan. 2017.

[53] A. Coluccia, A. Fascista, and G. Ricci, "A $k$-nearest neighbors approach to the design of radar detectors," *Signal Process.*, vol. 174, Sep. 2020, Art. no. 107609.

[54] W. Deng, H. Lei, and X. Zhou, "Traffic state estimation and uncertainty quantification based on heterogeneous data sources: A three detector approach," *Transp. Res. B, Methodol.*, vol. 57, pp. 132–157, Nov. 2013.

[55] Y. Gong, M. Abdel-Aty, and J. Park, "Evaluation and augmentation of traffic data including Bluetooth detection system on arterials," *J. Intell. Transp. Syst.*, pp. 1–13, 2019.

[56] M. Bernas, B. Płaczek, W. Korski, P. Loska, J. Smyła, and P. Szymała, "A survey and comparison of low-cost sensing technologies for road traffic monitoring," *Sensors*, vol. 18, no. 10, p. 3243, Sep. 2018.

[57] P. Zhang, L. Rui, X. Qiu, and R. Shi, "A new fusion structure model for real-time urban traffic state estimation by multisource traffic data fusion," in *Proc. 18th Asia–Pacific Netw. Oper. Manage. Symp. (APNOMS)*, Oct. 2016, pp. 1–6.

[58] F. Garcia, D. Martin, A. de la Escalera, and J. M. Armingol, "Sensor fusion methodology for vehicle detection," *IEEE Intell. Transp. Syst. Mag.*, vol. 9, no. 1, pp. 123–133, Jan. 2017.

[59] T. Adali, Y. Levin-Schwartz, and V. D. Calhoun, "Multimodal data fusion using source separation: Application to medical imaging," *Proc. IEEE*, vol. 103, no. 9, pp. 1494–1506, Sep. 2015.

[60] D. Lahat, T. Adali, and C. Jutten, "Multimodal data fusion: An overview of methods, challenges, and prospects," *Proc. IEEE*, vol. 103, no. 9, pp. 1449–1477, Sep. 2015.

[61] L. Zhang, Y. Xie, L. Xidao, and X. Zhang, "Multi-source heterogeneous data fusion," in *Proc. Int. Conf. Artif. Intell. Big Data*, May 2018, pp. 47–51.

[62] S. Jabbar, K. R. Malik, M. Ahmad, O. Aldabbas, M. Asif, S. Khalid, K. J. Han, and S. H. Ahmad, "A methodology real-time data fusion for localized big data analytics," *IEEE Access*, vol. 6, pp. 24510–24520, 2018.

[63] Y. Zheng, "Methodologies for cross-domain data fusion: An overview," *IEEE Trans. Big Data*, vol. 1, no. 1, pp. 16–34, Mar. 2015.

[64] A. Akbar, G. Kousiouris, H. Pervaiz, J. Sancho, P. Ta-Shma, F. Carrez, and K. Moessner, "Real-time probabilistic data fusion for large-scale IoT applications," *IEEE Access*, vol. 6, pp. 10015–10027, 2018.

[65] Y. Kawasaki, Y. Hara, and M. Kuwahara, "Traffic state estimation on a two-dimensional network by a state-space model," *Transp. Res. Procedia*, vol. 38, pp. 299–319, Jan. 2019.

[66] S. Liu, D. Zhang, and J. Li, "Study on traffic multi-source data fusion," *Int. J. Cognit. Informat. Natural Intell.*, vol. 13, no. 2, pp. 63–75, Apr. 2019.

[67] M. Rostami Shahrbabaki, A. A. Safavi, M. Papageorgiou, and I. Papamichail, "A data fusion approach for real-time traffic state estimation in urban signalized links," *Transp. Res. C, Emerg. Technol.*, vol. 92, pp. 525–548, Jul. 2018.

[68] L. Zhu, F. Guo, J. W. Polak, and R. Krishnan, "Multi-sensor fusion based on the data from bus GPS, mobile phone and loop detectors in travel time estimation," in *Proc. Transp. Res. Board 96th Annu. Meeting*, Washington, DC, USA, 2017, pp. 3417–3472.

[69] Y. Xia and X. Li, "Multi-sensor heterogeneous data representation for data-driven ITS," in *Proc. 16th Int. IEEE Conf. Intell. Transp. Syst. (ITSC)*, Oct. 2013, pp. 1750–1755.

[70] Z. Ning, X. Jianmin, L. PeiQun, and Z. Minjie, "An approach for real-time urban traffic state estimation by fusing multisource traffic data," in *Proc. 10th World Congr. Intell. Control Automat.*, Jul. 2012, pp. 4077–4081.

[71] E. Cipriani, S. Gori, and L. Mannini, "Traffic state estimation based on data fusion techniques," in *Proc. 15th Int. IEEE Conf. Intell. Transp. Syst.*, Sep. 2012, pp. 1477–1482.

[72] S. Mil and M. Piantanakulchai, "Travel time estimation based on fused traffic state data: Case studies in US and South Korea," *J. Eastern Asia Soc. Transp. Stud.*, vol. 11, pp. 1868–1884, Mar. 2015.

[73] T. Seo and T. Kusakabe, "Traffic state estimation using small imaging satellites and connected vehicles," *Transp. Res. Procedia*, vol. 34, pp. 4–11, Jan. 2018.

[74] Y. Guo and L. Yang, "Reliable estimation of urban link travel time using multi-sensor data fusion," *Information*, vol. 11, no. 5, p. 267, May 2020.

[75] X. Zhang, L. Xu, J. Li, and M. Ouyang, "Real-time estimation of vehicle mass and road grade based on multi-sensor data fusion," in *Proc. IEEE Vehicle Power Propuls. Conf. (VPPC)*, Oct. 2013, pp. 1–7.

[76] S. Khan, K. Dey, and M. Chowdhury, "Real-time traffic state estimation with connected vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 7, pp. 1687–1699, Mar. 2017.

[77] L. Zhu, F. Guo, J. W. Polak, and R. Krishnan, "Urban link travel time estimation using traffic states-based data fusion," *IET Intell. Transp. Syst.*, vol. 12, no. 7, pp. 651–663, Sep. 2018.

[78] X. Zhou, W. Wang, and L. Yu, "Traffic flow analysis and prediction based on gps data of floating cars," in *Proc. Int. Technol. Softw. Eng.*, 2012, pp. 497–508.

[79] T. Pamua and L. Kr, "The traffic flow prediction using Bayesian and neural networks," in *Proc. Conf. Intell. Transp. Syst.-Problems Perspect.*, 2016, pp. 105–126.

[80] R. Gravina, P. Alinia, H. Ghasemzadeh, and G. Fortino, "Multi-sensor fusion in body sensor networks: State-of-the-art and research challenges," *Inf. Fusion*, vol. 35, pp. 68–80, May 2017.

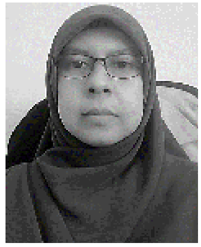
[81] A. Duivenvoorden, K. Lee, M. Raison, and S. Achiche, "Sensor fusion in upper limb area networks: A survey," in *Proc. Global Inf. Infrastructure Netw. Symp. (GIIS)*, Oct. 2017, pp. 56–63.

[82] A. Bris, N. Chehata, W. Ouerghemmi, C. Wendl, T. Postadjian, A. Puissant, and C. Mallet, "Decision fusion of remote-sensing data for land cover classification," in *Multimodal Scene Understanding*. New York, NY, USA: Academic, 2019, pp. 341–382.

[83] M. Muzammal, R. Talat, A. H. Sodhro, and S. Pirbhulal, "A multi-sensor data fusion enabled ensemble approach for medical data from body sensor networks," *Inf. Fusion*, vol. 53, pp. 155–164, Jan. 2020.

[84] E. Kanjo, E. M. G. Younis, and N. Sherkat, "Towards unravelling the relationship between on-body, environmental and emotion data using sensor information fusion approach," *Inf. Fusion*, vol. 40, pp. 18–31, Mar. 2018.
[85] L. Klein, L. Mihaylova, and E. Faouzi, "Sensor and data fusion: Taxonomy, challenges and applications," in *Handbook on Soft Computing for Video Surveillance*. Atlanta, GA, USA: Chapman & Hall, 2012, pp. 139–183.
[86] Y. Liu, S. He, B. Ran, and Y. Cheng, "A progressive extended Kalman filter method for freeway traffic state estimation integrating multi-source data," *Wireless Commun. Mobile Comput.*, vol. 2018, May 2018, Art. no. 6745726.
[87] J. Zhang, "Multi-source remote sensing data fusion: Status and trends," *Int. J. Image Data Fusion*, vol. 1, no. 1, pp. 5–24, Mar. 2010.
[88] R. Soua, A. Koesdwiady, and F. Karray, "Big-data-generated traffic flow prediction using deep learning and dempster-shafer theory," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2016, pp. 3195–3202.
[89] A. C. Nae and I. Dumitrache, "Fuzzy-logic adaptive control of traffic in an urban junction," in *Proc. 22nd Int. Conf. Control Syst. Comput. Sci.*, 2018, pp. 1–14.
[90] K. Bader, B. Lussier, and W. Schön, "A fault tolerant architecture for data fusion: A real application of Kalman filters for mobile robot localization," *Robot. Auton. Syst.*, vol. 88, pp. 11–23, Feb. 2017.
[91] S. He, H.-S. Shin, and A. Tsourdos, "Distributed joint probabilistic data association filter with hybrid fusion strategy," *IEEE Trans. Instrum. Meas.*, vol. 69 no. 1, pp. 286–300, Jan. 2020.
[92] M. Wielitzka, A. Busch, M. Dagen, T. Ortmaier, and G. Serra, "Unscented Kalman filter for state and parameter estimation in vehicle dynamics," in *Kalman Filters-Theory for Advanced Applications*. Rijeka, Croatia: InTech, 2018, pp. 56–75.
[93] E. Papapanagioutou, J. Kaths, and F. Busch, "Kalman filter for turning rate estimation at signalized intersections, based on Floating Car Data," in *Proc. Int. Sci. Conf. Mobility Transp. Urban Mobility-Shaping Future Together*, 2019, pp. 1–6.
[94] Y.-J. Byon, A. Shalaby, B. Abdulhai, C.-S. Cho, H. Yeo, and S. El-Tantawy, "Traffic condition monitoring with SCAAT Kalman filter-based data fusion in Toronto, Canada," *KSCE J. Civil Eng.*, vol. 23, no. 2, pp. 810–820, Feb. 2019.
[95] H. Zhang and A. Poschinger, "Estimation of turning ratio at intersections based on detector data using Kalman filter," *Transp. Res. Procedia*, vol. 41, pp. 673–687, 2019.
[96] F. Ottaviano, F. Cui, and A. H. F. Chow, "Modeling and data fusion of dynamic highway traffic," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2644, no. 1, pp. 92–99, Jan. 2017.
[97] M. Saeedmanesh, A. Geroliminis, and L. Nikolas, "A real-time state estimation approach for multi-region MFD traffic systems based on extended Kalman filter," in *Proc. Annu. Meeting Transp. Res. Board*, 2019, pp. 2756.
[98] X. Li and Z. Guo, "Multi-source information fusion model of traffic lifeline based on improved DS evidence theory," in *Proc. 26th Int. Conf. Geoinformatics*, Jun. 2018, pp. 1–6.
[99] X. Du, M. El-Khamy, J. Lee, and L. Davis, "Fused DNN: A deep neural network fusion approach to fast and robust pedestrian detection," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2017, pp. 953–961.
[100] A. Essien, I. S. P. Petrounias, and S. Sampaio, "Improving urban traffic speed prediction using data source fusion and deep learning," in *Proc. IEEE Int. Conf. Big Data Smart Comput.*, Feb. 2019, pp. 1–8.
[101] D. Velusamy and G. K. Pugalendhi, "Fuzzy integrated Bayesian Dempster–Shafer theory to defend cross-layer heterogeneity attacks in communication network of smart grid," *Inf. Sci.*, vol. 479, pp. 542–566, Apr. 2019.
[102] O. A. Wahab, H. Otrok, and A. Mourad, "A cooperative watchdog model based on Dempster–Shafer for detecting misbehaving vehicles," *Comput. Commun.*, vol. 41, pp. 43–54, Mar. 2014.
[103] S. A. Mostafa, R. Darman, S. H. Khaleefah, A. Mustapha, N. Abdullah, and H. Hafit, "A general framework for formulating adjustable autonomy of multi-agent systems by fuzzy logic," in *Proc. KES Int. Symp. Agent Multi-Agent Syst., Technol. Appl.* Cham, Switzerland: Springer, Jun. 2018, pp. 23–33.
[104] B. Khaleghi, A. Khamis, F. O. Karray, and S. N. Razavi, "Multisensor data fusion: A review of the state-of-the-art," *Inf. Fusion*, vol. 14, no. 1, pp. 28–44, Jan. 2013.
[105] M. Balta and İ. Özçelik, "A 3-stage fuzzy-decision tree model for traffic signal optimization in urban city via a SDN based VANET architecture," *Future Gener. Comput. Syst.*, vol. 104, pp. 142–158, Mar. 2020.
[106] W. Chen, J. An, L. Fu, G. Xie, M. Z. Bhuiyan, and K. Li, "A novel fuzzy deep-learning approach to traffic flow prediction with uncertain spatial–temporal data features," *Future Gener. Comput. Syst.*, vol. 89, pp. 78–88, Dec. 218.
[107] P.-W. Wang, H.-B. Yu, L. Xiao, and L. Wang, "Online traffic condition evaluation method for connected vehicles based on multisource data fusion," *J. Sensors*, vol. 2017, Aug. 2017, Art. no. 7248189.
[108] Z. Bouyahia, H. Haddad, N. Jabeur, and S. Derrode, "Real-time traffic data smoothing from GPS sparse measures using fuzzy switching linear models," *Procedia Comput. Sci.*, vol. 110, pp. 143–150, Jan. 2017.
[109] S. H. Rezatofighi, A. Milan, Z. Zhang, Q. Shi, A. Dick, and I. Reid, "Joint probabilistic data association revisited," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 3047–3055.
[110] F. García, A. de la Escalera, and J. M. Armingol, "Joint probabilistic data association fusion approach for pedestrian detection," in *Proc. IEEE Intell. Vehicles Symp.*, Jun. 2013, pp. 1344–1349.
[111] Y. Hu, W. Zhan, and M. Tomizuka, "A framework for probabilistic generic traffic scene prediction," in *Proc. 21st Int. Conf. Intell. Transp. Syst.*, Nov. 2018, pp. 2790–2796.
[112] Z. Zhu, B. Peng, C. Xiong, and L. Zhang, "Short-term traffic flow prediction with linear conditional Gaussian Bayesian network," *J. Adv. Transp.*, vol. 50, no. 6, pp. 1111–1123, 2016.
[113] D. Lam, M. Wei, and D. Wunsch, "Clustering data of mixed categorical and numerical type with unsupervised feature learning," *IEEE Access*, vol. 3, pp. 1605–1613, 2015.
[114] B. Yu, X. Song, F. Guan, Z. Yang, and B. Yao, "$K$-nearest neighbor model for multiple-time-step prediction of short-term traffic condition," *J. Transp. Eng.*, vol. 142, no. 6, Jun. 2016, Art. no. 04016018.
[115] C. Gong, Z.-G. Su, P.-H. Wang, and Q. Wang, "Cumulative belief peaks evidential $k$-nearest neighbor clustering," *Knowl.-Based Syst.*, vol. 200, Jul. 2020, Art. no. 105982.
[116] H. Yu, N. Ji, Y. Ren, and C. Yang, "A special event-based $k$-nearest neighbor model for short-term traffic state prediction," *IEEE Access*, vol. 7, pp. 81717–81729, 2019.
[117] D. Draskovic, V. Gencel, S. Zitnik, M. Bajec, and B. Nikolic, "A software agent for social networks using natural language processing techniques," in *Proc. 24th Telecommun. Forum*, Nov. 2016, pp. 1–4.
[118] A. Jamshidnejad, B. De Schutter, and M. J. Mahjoob, "Urban traffic control using a fuzzy multi-agent system," in *Proc. Eur. Control Conf. (ECC)*, Jul. 2015, pp. 3041–3046.
[119] E. B. Boussada, M. B. Ayed, and A. M. Alimi, "Data fusion classification method based on multi agents system," in *Proc. Int. Conf. Intell. Syst. Design Appl.*, 2017, pp. 863–870.
[120] H. Hamidi and A. Kamankesh, "An approach to intelligent traffic management system using a multi-agent system," *Int. J. Intell. Transp. Syst. Res.*, vol. 16, no. 2, pp. 112–124, May 2018.
[121] M. K. Tan, H. S. E. Chuo, R. K. Y. Chin, K. B. Yeo, and K. T. K. Teo, "Hierarchical multi-agent system in traffic network signalization with improved genetic algorithm," in *Proc. IEEE Int. Conf. Artif. Intell. Eng. Technol. (IICAIET)*, Nov. 2018, pp. 1–6.
[122] M. Xu, K. An, L. H. Vu, Z. Ye, J. Feng, and E. Chen, "Optimizing multi-agent based urban signal traffic signal control system," *J. Intell. Transp. Syst.*, vol. 23, no. 4, pp. 357–369, 2018.
[123] M. Panicker, T. Mitha, K. Oak, A. M. Deshpande, and C. Ganguly, "Multisensor data fusion for an autonomous ground vehicle," in *Proc. Conf. Adv. Signal Process. (CASP)*, Jun. 2016, pp. 507–512.
[124] H. Lu, Y. Li, M. Chen, H. Kim, and S. Serikawa, "Brain intelligence: Go beyond artificial intelligence," *Mobile Netw. Appl.*, vol. 23, no. 2, pp. 368–375, 2017.
[125] A. Koesdwiady, R. Soua, and F. Karray, "Improving traffic flow prediction with weather information in connected cars: A deep learning approach," *IEEE Trans. Veh. Technol.*, vol. 65, no. 12, pp. 9508–9517, Dec. 2016.
[126] S. Ding, C. Su, and J. Yu, "An optimizing BP neural network algorithm based on genetic algorithm," *Artif. Intell. Rev.*, vol. 36, no. 2, pp. 153–162, Aug. 2011.
[127] N. Alqudah and Q. Yaseen, "Machine learning for traffic analysis: A review," *Procedia Comput. Sci.*, vol. 170, pp. 911–916, Dec. 2020.
[128] L. Li, B. Du, Y. Wang, L. Qin, and H. Tan, "Estimation of missing values in heterogeneous traffic data: Application of multimodal deep learning model," *Knowl.-Based Syst.*, vol. 194, Apr. 2020, Art. no. 105592.
[129] R. Nowak, R. Biedrzycki, and J. Misiurewicz, "Machine learning methods in data fusion systems," in *Proc. 13th Int. Radar Symp.*, May 2012, pp. 400–405.

[130] Y. Gu, W. Lu, X. Xu, L. Qin, Z. Shao, and H. Zhang, "An improved Bayesian combination model for short-term traffic prediction with deep learning," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 3, pp. 1332–1342, Mar. 2019.
[131] S. Din, A. Ahmad, A. Paul, M. M. U. Rathore, and G. Jeon, "A cluster-based data fusion technique to analyze big data in wireless multi-sensor system," *IEEE Access*, vol. 5, pp. 5069–5083, 2017.
[132] N. Dogru and A. Subasi, "Traffic accident detection using random forest classifier," in *Proc. 15th Learn. Technol. Conf.*, Feb. 2018, pp. 40–45.
[133] X. Shen and Y. Zhu, "General heterogeneous sensor estimation fusion-system fusion method," in *Proc. 36th Chin. Control Conf. (CCC)*, Jul. 2017, pp. 5173–5178.
[134] P. Cai, Y. Wang, G. Lu, P. Chen, c. Ding, and J. Sun, "A spatiotemporal correlative $k$-nearest neighbor model for short-term traffic multistep forecasting," *Transp. Res. C, Emerg. Technol.*, vol. 62, pp. 21–34, Jan. 2016.
[135] E. Chong, C. Han, and F. C. Park, "Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies," *Expert Syst. Appl.*, vol. 83, pp. 187–205, Oct. 2017.
[136] T. Chai and R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)?—Arguments against avoiding RMSE in the literature," *Geosci. Model. Develop.*, vol. 7, no. 3, pp. 1247–1250, Jun. 2014.
[137] A. de Myttenaere, B. Golden, B. Le Grand, and F. Rossi, "Mean absolute percentage error for regression models," *Neurocomputing*, vol. 192, pp. 38–48, Jun. 2016.
[138] J. A. Kamińska, "A random forest partition model for predicting $NO_2$ concentrations from traffic flow and meteorological conditions," *Sci. Total Environ.*, vol. 651, pp. 475–483, Feb. 2019.
[139] T. Zhou, D. Jiang, Z. Lin, G. Han, X. Xu, and J. Qin, "Hybrid dual Kalman filtering model for short-term traffic flow forecasting," *IET Intell. Transp. Syst.*, vol. 13, no. 6, pp. 1023–1032, Jun. 2019.
[140] S. Kim and H. Kim, "A new metric of absolute percentage error for intermittent demand forecasts," *Int. J. Forecasting*, vol. 32, no. 3, pp. 669–679, Jul. 2016.
[141] B. Ran, H. Tan, J. Feng, W. Wang, Y. Cheng, and P. Jin, "Estimating missing traffic volume using low multilinear rank tensor completion," *J. Intell. Transp. Syst.*, vol. 20, no. 2, pp. 152–161, Mar. 2016.
[142] A. Ghosh, R. Sharma, and P. K. Joshi, "Random forest classification of urban landscape using landsat archive and ancillary data: Combining seasonal maps with decision level fusion," *Appl. Geography*, vol. 48, pp. 31–41, Mar. 2014.
[143] J. Steenbruggen, E. Tranos, and P. Rietveld, "Traffic incidents in motorways: An empirical proposal for incident detection using data from mobile phone operators," *J. Transp. Geography*, vol. 54, pp. 81–90, Jun. 2016.
[144] M. Rostami-Shahrbabaki, A. A. Safavi, M. Papageorgiou, P. Setoodeh, and I. Papamichail, "State estimation in urban traffic networks: A two-layer approach," *Transp. Res. C, Emerg. Technol.*, vol. 115, Jun. 2020, Art. no. 102616.
[145] T.-H. Chang, A. Y. Chen, C.-W. Chang, and C.-H. Chueh, "Traffic speed estimation through data fusion from heterogeneous sources for first response deployment," *J. Comput. Civil Eng.*, vol. 28, no. 6, Nov. 2014, Art. no. 04014018.
[146] L. Li, X. Sheng, B. Du, Y. Wang, and B. Ran, "A deep fusion model based on restricted Boltzmann machines for traffic accident duration prediction," *Eng. Appl. Artif. Intell.*, vol. 93, Aug. 2020, Art. no. 103686.
[147] A. Essien, I. Petrounias, P. Sampaio, and S. Sampaio, "A deep-learning model for urban traffic flow prediction with traffic events mined from Twitter," *World Wide Web*, vol. 2020, pp. 1–14, Mar. 2020.
[148] Y. Lin and R. Li, "Real-time traffic accidents post-impact prediction: Based on crowdsourcing data," *Accident Anal. Prevention*, vol. 145, Sep. 2020, Art. no. 105696.
[149] H. Zhu, H. Leung, and K.-V. Yuen, "A joint data association, registration, and fusion approach for distributed tracking," *Inf. Sci.*, vol. 324, pp. 186–196, Dec. 2015.
[150] G. Bello-Orgaz, J. Jung, and D. Camaho, "Social big data: Recent achievements and new challenges," *Inf. Fusion*, vol. 28, pp. 45–49, Mar. 2016.
[151] M. Bouain, D. Berdjag, and R. B. Atitallah, "Exploring high-level synthesis tools for vehicle perception tasks," in *Proc. 9th Eur. Congr. Embedded Real Time Softw. Syst.*, 2018, pp. 1–5.
[152] Z. Liu, W. Zhang, S. Lin, and T. Q. S. Quek, "Heterogeneous sensor data fusion by deep multimodal encoding," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 3, pp. 479–491, Apr. 2017.
[153] S. Elkosantini and A. Frikha, "Decision fusion for signalized intersection control," *Kybernetes*, vol. 44, no. 1, pp. 57–76, Jan. 2015.
[154] S. P. Singh, A. Sharma, and R. Kumar, "Design and exploration of load balancers for fog computing using fuzzy logic," *Simul. Model. Pract. Theory*, vol. 101, May 2020, Art. no. 102017.
[155] B. Alkouz and Z. Al Aghbari, "SNSJam: Road traffic analysis and prediction by fusing data from multiple social networks," *Inf. Process. Manage.*, vol. 57, no. 1, Jan. 2020, Art. no. 102139.
[156] L. Li, R. Jiang, Z. He, X. Chen, and X. Zhou, "Trajectory data-based traffic flow studies: A revisit," *Transp. Res. C, Emerg. Technol.*, vol. 114, pp. 225–240, May 2020.
[157] Q. Mei, M. Gül, and N. Shirzad-Ghaleroudkhani, "Towards smart cities: Crowdsensing-based monitoring of transportation infrastructure using in-traffic vehicles," *J. Civil Struct. Health Monitor.*, vol. 10, no. 4, pp. 653–665, Sep. 2020.
[158] H. Mai-Tan, H.-N. Pham-Nguyen, N. X. Long, and Q. T. Minh, "Mining urban traffic condition from crowd-sourced data," *Social Netw. Comput. Sci.*, vol. 1, no. 4, pp. 1–16, Jul. 2020.



**SHAFIZA ARIFFIN KASHINATH** received the B.Sc. degree in computer science from Universiti Teknologi Malaysia (UTM), in 2004. She is currently pursuing the M.Sc. degree in information technology with Universiti Tun Hussein Onn Malaysia (UTHM). She is also a Software Developer with Sena Traffic Systems Sdn. Bhd., Malaysia. Her research interests include the areas of traffic flow analysis and artificial intelligence.

**SALAMA A. MOSTAFA** received the B.Sc. degree in computer science from the University of Mosul, Iraq, in 2003, and the M.Sc. and Ph.D. degrees in information and communication technology from Universiti Tenaga Nasional (UNITEN), Malaysia, in 2011 and 2016, respectively. He is currently the Head of the Center of Intelligent and Autonomous Systems (CIAS), Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia (UTHM). His specialization and research interests include the areas of autonomous agent, human–computer collaboration, machine learning, optimization, and software quality assurance.

**AIDA MUSTAPHA** (Member, IEEE) received the B.Sc. degree in computer science from Michigan Technological University, in 1998, the M.Sc. degree in computer science from UKM, Malaysia, in 2004, and the Ph.D. degree in artificial intelligence with a focus on dialogue systems. She is currently an Active Researcher in the area of computational linguistics, soft computing, data mining, and agent-based systems.

**HAIRULNIZAM MAHDIN** received the Ph.D. degree in information technology from Deakin University, Australia, in 2012. He is currently the Deputy Dean of the Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia (UTHM). His research interests include data management, the IoT, RFID, software engineering, and web technology.

**DAVID LIM** is currently the Deputy General Manager with the Engineering Research and Development Department, Sena Traffic Systems Sdn. Bhd.

**MOAMIN A. MAHMOUD** received the bachelor's degree in mathematics from the College of Mathematics and Computer Science, University of Mosul, Iraq, in 2007, the Master of Information Technology degree from the College of Graduate Studies, Universiti Tenaga Nasional (UNITEN), Malaysia, in 2010, and the Ph.D. degree in information and communication technology from UNITEN, in 2013. Since 2014, he has been with the Department of Software Engineering, Universiti Tenaga Nasional, as a Senior Lecturer. His current research interests include artificial intelligence, distributed and autonomous systems, complex adaptive systems, and the IoT-based smart systems.

**MAZIN ABED MOHAMMED** received the B.Sc. degree in computer science from the University of Anbar, Iraq, in 2008, the M.Sc. degree in information technology from UNITEN, Malaysia, in 2011, and the Ph.D. degree in information technology from UTeM, Malaysia, in 2019. He is currently a Lecturer with the College of Computer Science and Information Technology, University of Anbar. His research interests include artificial intelligence, biomedical computing, and optimization.

**BANDER ALI SALEH AL-RIMY** received the B.Sc. degree in computer engineering from the Faculty of Engineering, Sana'a University, Yemen, in 2003, the M.Sc. degree in information technology from OUM, Malaysia, in 2013, and the Ph.D. degree in computer science from the Faculty of Engineering, Universiti Teknologi Malaysia (UTM), Johor Bahru, Malaysia, in 2019, with a focus on information security. He is currently a Senior Lecturer with UTM. His research interests include but not limited to Malware, IDS, network security, and routing technologies. He was a recipient of several academic awards and recognitions, including but not limited to the UTM Alumni Award, the UTM Best Postgraduate Student Award, the UTM Merit Award, the UTM Excellence Award, the OUM Distinction Award, and the Best Research Paper Award.

**MOHD FARHAN MD FUDZEE** (Senior Member, IEEE) received the Diploma degree in computer science from Universiti Teknologi Malaysia, the Bachelor of Science degree (Hons.) in information technology, the Master of Science degree in information technology from Universiti Teknologi MARA, in 2001, 2003, and 2005, respectively, and the Doctor of Philosophy degree from Deakin University, Australia, in 2012. He is currently working with the Faculty of Computer Science and Information Technology (FSKTM), Universiti Tun Hussein Onn Malaysia (UTHM). He is also the Chief Information Officer (CIO) of UTHM. His research interests include computing, such as content adaptation and energy-related solution, multimedia security, decision making algorithms, web technology, e-government, and creative applications.

**TAN JHON YANG** is currently the Manager of the urban traffic system with the Engineering Research and Development Department, Sena Traffic Systems Sdn. Bhd.