WILEY | Hindawi

Review Article

A Review of Traffic Congestion Prediction Using Artificial Intelligence

Mahmuda Akhtar (ID) **and Sara Moridpour** (ID)

Department of Civil and Infrastructure Engineering, RMIT University, Melbourne, VIC 3000, Australia

Correspondence should be addressed to Mahmuda Akhtar; s3799862@student.rmit.edu.au

In recent years, traffic congestion prediction has led to a growing research area, especially of machine learning of artificial intelligence (AI). With the introduction of big data by stationary sensors or probe vehicle data and the development of new AI models in the last few decades, this research area has expanded extensively. Traffic congestion prediction, especially short-term traffic congestion prediction is made by evaluating different traffic parameters. Most of the researches focus on historical data in forecasting traffic congestion. However, a few articles made real-time traffic congestion prediction. This paper systematically summarises the existing research conducted by applying the various methodologies of AI, notably different machine learning models. The paper accumulates the models under respective branches of AI, and the strength and weaknesses of the models are summarised.

1. Introduction

Artificial intelligence (AI) is the most important branch of computer science in this era of big data. AI was born 50 years ago and came a long way, making encouraging progress, especially in machine learning, data mining, computer vision, expert systems, natural language processing, robotics, and related applications [1]. Machine learning is the most popular branch of AI. Other classes of AI include probabilistic models, deep learning, artificial neural network systems, and game theory. These classes are developed and applied in a wide range of sectors. Recently, it has been the leading research area in transportation engineering, especially in traffic congestion prediction.

Traffic congestion has a direct and indirect impact on a country's economy and its dwellers' health. According to Ali et al. [2], traffic congestion causes Pak Rs. 1 million every day in terms of opportunity cost and fuel consumption due to traffic congestion. Traffic congestion affects on individual level as well. Time loss, especially during peak hours, mental stress, and the added pollution to the global warming are also some important factors caused due to traffic congestion.

Ensuring economic growth and the road users' comfort are the two requirements for the development of a country, which is impossible without smooth traffic flow. With the development in the transportation sector by collecting traffic information, authorities are putting more attention on traffic congestion monitoring. Traffic congestion prediction provides the authorities with the required time to plan in the allocation of resources to make the journey smooth for travellers. Traffic congestion prediction problem discussed in this paper can be defined as an estimation of parameters related to traffic congestion into the short-term future, e.g., 15 minutes to a few hours by applying different AI methodologies by using collected traffic data. There are usually five parameters to evaluate, including traffic volume, traffic density, occupancy, traffic congestion index, and travel time while monitoring and predicting traffic congestions. Depending on the nature of the collected data, a variety of AI approaches are applied to evaluate the congestion parameters. This article systematically discusses the models and their advantage and disadvantages. The primary motivation of this review is to gather the articles focusing solely on traffic congestion prediction models. The keywords used in

the search process included "traffic congestion prediction" OR "traffic congestion estimation" OR "congestion prediction modelling" OR "prediction of traffic congestion" OR "road congestion forecast" OR "traffic congestion forecast." For efficient screening, research paper search was done according to year using search engines like Scopus, Google Scholar, and Science Direct. After collecting all the peer-reviewed journal and conference papers written in the English language, 48 articles were found for review. Any studies focusing on the cause of traffic congestion, traffic congestion control, traffic congestion impact, traffic congestion propagation, traffic congestion prevention, etc. were excluded from this manuscript.

A general layout of the prediction approaches is provided in Section 2. The data collection sources and congestion forecasting models are explained in Sections 3–6 and they provide the overall discussion and concluding remarks.

2. General Layout

Traffic congestion forecasting has two basic steps of data collection and prediction model development. Every step of the methodology is important and may affect the results if not done correctly. After data collection, data processing plays a vital role to prepare the training and testing datasets. Case area differs for different research. After developing the model, it is validated with other base models and ground true results. Figure 1 shows the general components of traffic congestion prediction studies. These branches were further divided into more specific sub-branches and are discussed in the following sections.

3. Data Source

Traffic datasets used in different studies can be mainly divided into two classes, including stationary and probe data. Stationary data can be further divided into sensor data and fixed cameras. On the other hand, probe data that were used in the studies were GPS data mounted on vehicles.

Stationary sensors continuously capture spatiotemporal data of traffic. However, sensor operation may interrupt anytime. Authorities should always consider this temporary failure of the sensor while planning by using this data. The advantage of the sensor data is that there is no confusion on the location of the vehicles. The most used dataset was Performance Measurement System (PeMS) that collects highway data across all major metropolitan areas of the State of California of traffic flow, sensor occupancy, and travel speed in real-time. Most of the studies used dataset from the I-5 highway, in San Diego, California, every 5 minutes [3–6]. Other systems included the Genetec blufaxcloud travel-time system engine (GBTTSE) [7] and the Topologically Integrated Geographic Encoding and Referencing (TIGER) line graph [8].

On the other hand, probe data has the advantage of covering the entire road network. A network consists of different structured roads. Therefore, studies, especially those that considered the network wide area, used probe data. The most used dataset was GPS data collecting every

second from approximately 20000 taxies of Beijing, China. Data included the taxi number, the latitude-longitude of the vehicle, timestamp when sampling, and whether there was a passenger or not. Data updating frequency of this dataset varies from 10 s to 5 min according to the quality of GPS device [4, 5, 9]. Other probe data included low-frequency Probe Vehicle Data (PVD) [10] and bus GPS data [11, 12]. However, sometimes probe data show significant fluctuation. Besides, map matching is usually a must for probe data. But data can minimize this limitation. Probe data collected from one city cannot be used directly for modelling other city networks. This is because the data collected from Beijing, China, includes latitude-longitude of the vehicle, which is unique. However, a generalised model using probe data can be generated for different cities.

Other data sources, e.g., data from tolling system and data provided by transportation authority, will add more reliable data as the sources are dependable. However, a lot of the times, study area needs to be adjusted as in most cases, tolled road information is not available. Tracking cellular phone movements without privacy breach can also be a source of data. However, the heterogeneity of the vehicle distribution will be hard to determine from this dataset, if not impossible. Besides, due to pedestrian or cyclists travelling through the sidewalk, there might be many outliers in the dataset if modelling is done for a road network. Data collected from a questionnaire to the general public/drivers may provide a misleading result [13].

3.1. Clustering Algorithms. Some studies use clustering the acquired data before applying the main congestion models of prediction. This hybrid modelling technique is applied to fine-tune the input values and to use them in the training phase. Figure 2 shows the commonly used AI clustering models in this field of research. The models are described briefly in this section.

Fuzzy C-Means (FCM) is a popular nondeterministic clustering technique in data mining. In traffic engineering researches, traffic pattern recognition plays an important role. Besides, these studies often face the limitation of missing or incomplete data. To deal with these constraints, FCM has become a commonly applied clustering technique. The advantage of this approach is, unlike original C-means clustering methods, it can overcome the issue of getting trapped in the local optimum [14]. However, FCM requires setting a predefined cluster number, which is not always possible while dealing with massive data without any prior knowledge of the data dimension. Besides, this model becomes computationally expensive with data size increment. Different studies have applied FCM successfully by improving its limitations. Some studies changed the fuzzy index value for each FCM algorithm execution [15], some calculated the Davies-Bouldin (DB) index [10], while others applied the K-means clustering algorithm [16, 17].

K-means clustering is an effective and relatively flexible algorithm while dealing with large datasets. It is a popular unsupervised machine learning algorithm. Depending on the features, cluster number varied from two [18] to 50
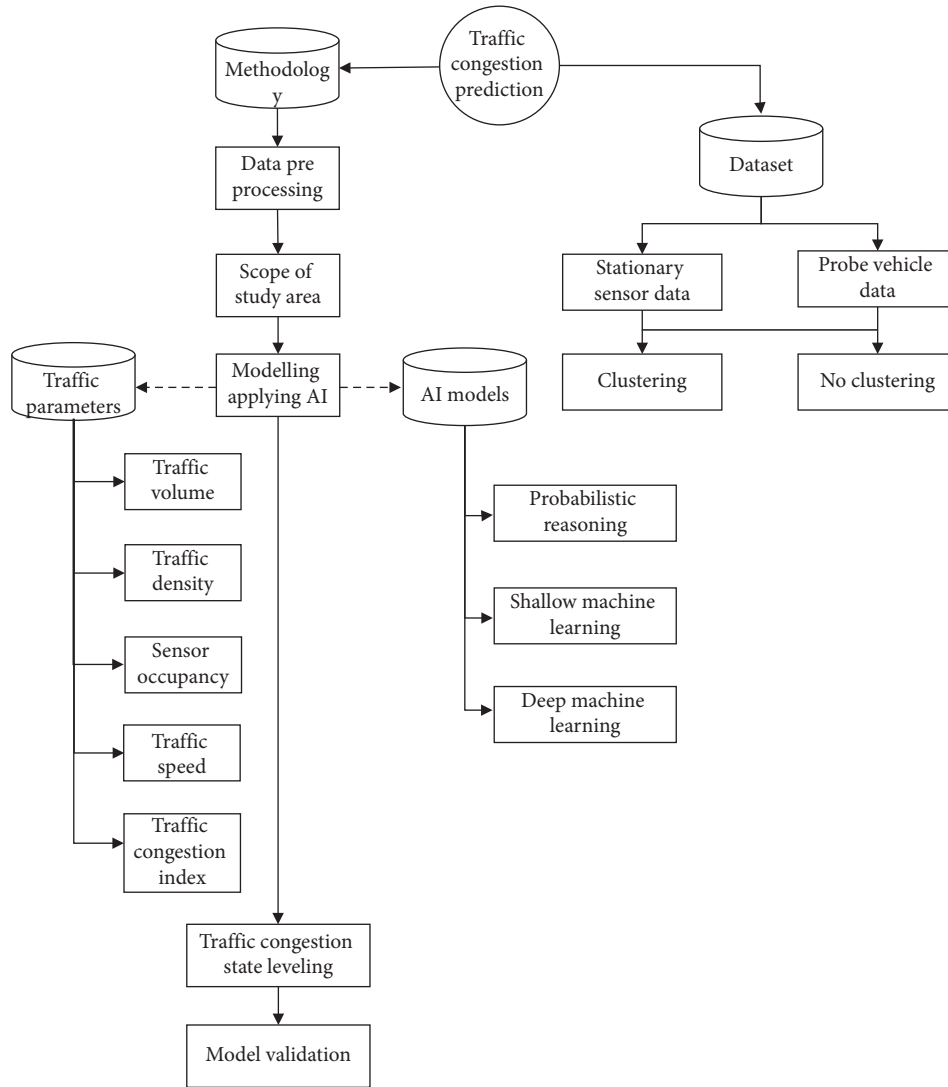
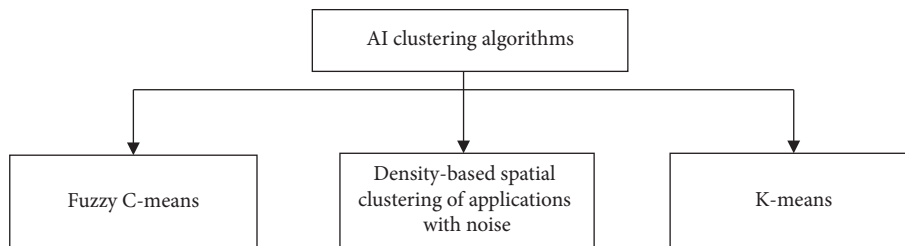FIGURE 1: The layout of traffic congestion prediction system.

FIGURE 2: Commonly used AI clustering algorithms.

[19–21]. Like FCM, K-means clustering requires a predefined cluster number and selecting K original cluster centres. GAP [22] and WEKA toolbox [23] were used to define the value. For large datasets, as the sample distribution is unknown in the beginning, it is not always possible to fulfil these two requirements. A few studies used adaptive K-means clustering overcoming the limitations and exploited the pattern using principal component analysis (PCA) [24, 25].

DBSCAN is more of a general clustering application in machine learning and data mining. This method overcomes the limitation of FCM of predefining the cluster number. It

can automatically generate arbitrary cluster shapes surrounded by clusters of different characteristics and can easily recognise outlier. However, it requires two parameters to preset. A suitable parameter determination method, e.g., trial and error method [8] and human judgement [26] makes the model computationally expensive and requires a clear understanding of the dataset.

From the above discussion, it is concluded that only 16 out of 48 studies have done clustering before applying prediction models. Several time-series models and shallow machine learning (SML) algorithms have used clustering approach. However, deep learning algorithms can process input data on different layers of the model, thus may not need clustering beforehand.

4. Applied Methodology

Traffic flow is a complex amalgamation of heterogenous traffic fleet. Thus, traffic pattern prediction modelling could be an easy and efficient congestion prediction approach. However, depending on the data characteristics and quality, different classes of AI are applied in various studies. Figure 3 shows the main branches—probabilistic reasoning and machine learning (ML). Machine learning comprised of both shallow and deep learning algorithms. However, with the progress of this article, these sections were subdivided into detailed algorithms.

To generalise traffic congestion forecasting studies using different models is not straight forward. The common factors of all the articles include the study area, data collection horizon, predicted parameter, prediction intervals, and validation procedure. Most of the articles took studied corridor segment as the study area [5, 27–30]. Other study areas included the traffic network [31, 32], ring road [9], and arterial road [33]. Data collection horizon varied from 2 years [34] to less than a day [35] in the studies. Congestion estimation is done predicting traffic flow parameters, e.g., traffic speed [4], density, speed [5], and congestion index [31], to mention a few. The Congestion Index (CI) approach is suitable to monitor the congestion level continuously in a spatiotemporal dimension. Studies those compared their results with the ground truth value or with other models used mean absolute error (MAE) (equation (1)), symmetric mean absolute percentage error (sMAPE) (equation (1)), MAPE, root-mean-squared error (RMSE) (equation (3)), false positive rate (FPR) (equation (4)), and detection rate (DR) (equation (5)). Many studies used SUMO to validate their models:

$$MAE = \frac{1}{n} \sum_{i=1}^{n} \left| Y_i - \overline{Y_i} \right|, \tag{1}$$

$$sMAPE = \frac{1}{n} \sum_{i=1}^{n} \frac{\left| \overline{Y}_i - Y_i \right|}{\left(\left| \overline{Y}_i \right| + \left| Y_i \right| \right)/2} \cdot 100, \tag{2}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} \left(\overline{Y}_i - Y_i \right)^2}{n}}, \tag{3}$$
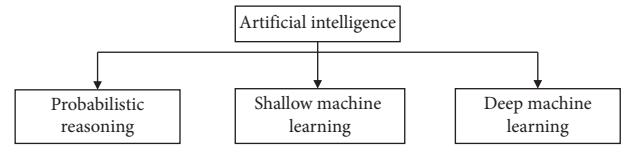

Figure 3: Branches of artificial intelligence in this article.

where $\overline{Y}$ = original value, Y_i = predicted value, and n = number of instances.

$$FPR = \frac{FP}{TN + FP}, \tag{4}$$

$$DR = \frac{TP}{FN + TP}, \tag{5}$$

where FP, TN, FN, and TP represent the false positive, true negative, false negative, and true positive, respectively.

The rest of this section will discuss the methodology the authors have applied in the studies.

4.1. Probabilistic Reasoning. Probabilistic reasoning is a significant section of AI. It is applied to deal with the field of uncertain knowledge and reasoning. A variety of these algorithms are commonly used in traffic congestion prediction studies. The studies discussed hereunder probabilistic reasoning is shown in Figure 4.

4.1.1. Fuzzy Logic. Zadeh is a commonly applied model in dynamic traffic congestion prediction as it allows vagueness instead of binary outcomes. In this method, several membership functions are developed those represent the degree of truth. With the vastness with time, traffic data are becoming complex and nonlinear. Due to its ability to deal with uncertainty in the dataset, fuzzy logic has become popular in traffic congestion prediction studies.

A fuzzy system comprises of several fuzzy sets, which is built of membership functions. There are usually three codification shapes to choose for the membership functions (MFs) of input: triangular, trapezoidal, and Gauss function. The fuzzy rule-based system (FRBS) is the most common fuzzy logic system in traffic engineering research. It consists of several IF-THEN rules that logically relate the input variables with output. It can effectively deal with the complexity resulting from real-world traffic situations by representing them in simple rules. These rules combine the relations among different traffic states to detect the resulting traffic condition [36]. However, with the growth in data complexity, the total number of rules also grows, lessening the accuracy of the whole system, thus making it computationally expensive. To better manage this problem, two types of fuzzy logic controls are applied. In hierarchical control (HFRBS), according to the significance, the input variables are ordered and MFs are employed. Figure 5 shows a simple HFRBS structure. MFs are optimized by applying different algorithms, e.g., genetic algorithm (GA) [30], hybrid genetic algorithm (GA), and cross-entropy (CE)
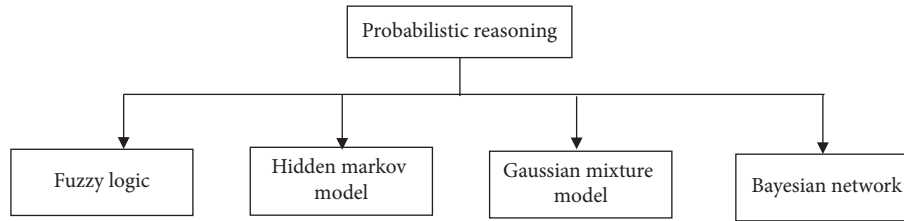
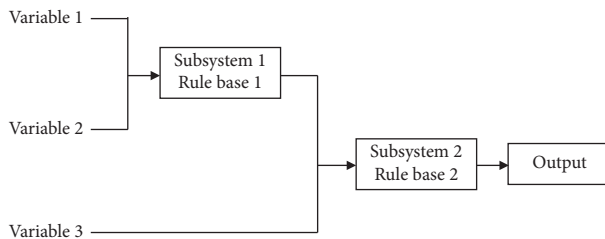Figure 4: Subdivision of probabilistic reasoning models.

Figure 5: A simple structure of HFRBS.

[28, 37] compared the performance of evolutionary crisp rule learning (ECRL) and evolutionary fuzzy rule learning (EFRL) for road traffic congestion prediction. It was seen that ECRL models outperformed EFRL in terms of averaged accuracy and no of rules but was computationally expensive.

The Takagi-Sugeno-Kang (TSK) (FRBS) model is one of the simple fuzzy models due to its mathematical treatability. A weighted average computes the output of this model. Another simple FRBS model is Mamdani-type model. The output of this model is a fuzzy set which needs defuzzification, which is time-consuming. Due to its good interpretability, it can improve the accuracy of fuzzy linguistic models. Cao and Wang [3] applied this model to show the congestion severity change among road grades. A few studies used this method to fuse heterogenous parameters [7, 13]. The TSK model works on improving the interpretability of an accurate fuzzy model. TSK is applied for its fast calculation characteristics [37].

The fuzzy comprehensive evaluation (FCE) uses the principle of fuzzy transformation and maximum membership degree. This model consists of several layers, which is a useful objective evaluation method, assessing all relevant factors. The number of layers depends on the objective complicacy and the number of factors. Kong et al. [4] and Yang et al. [5] applied FCE in which the weights and the fuzzy matrix of multi-indexes were adapted according to the traffic flow to estimate traffic congestion state. Adaptive control adjusts weight coefficient based on judgement matrix. Certain weights are assigned to calculate the membership degree of the parameters [35].

Other than GA and PSO, Ant Colony Optimization (ACO) algorithm was also introduced by Daissaoui et al. [38] in fuzzy logic system. They provided the theory for a smart city, where each vehicle GPS data was taken as a pheromone, consistent with the concept of ACO. The objective was to predict traffic congestion one minute ahead from the information (pheromone) provided by past cars. However, the article does not give any result on support to the model.

As discussed before, with the development of optimisation algorithms, optimisation of the fuzzy logic system's membership functions is becoming diverse. With time, the simplest form of FRBS-TSK has become popular due to its good interpretability. Some other sectors of transportation where fuzzy logic models are popular include traffic light/signal control [39, 40], traffic flow prediction (Zhang and Ye [41]), traffic accident prediction [42], and modified fuzzy logic for freeway travel time estimation (Zhang and Ge [43]).

The fuzzy logic system is the only probabilistic reasoning model that can have an outcome of more than congested/noncongested state of the traffic state. This is one of the main advantages that has made this methodology popular. However, no study has provided any reasonable logic on selecting the membership function, which is a significant limitation of fuzzy logic models.

4.1.2. Hidden Markov Model. The hidden Markov model (HMM) is a combination of stochastic characteristics of Markov process and discrete characteristics of Markov chains. It is a stochastic, time-series event recognition technique. Some studies have applied Markov chain model for traffic pattern recognition during congestion prediction [21, 25, 44]. Pearson correlation coefficient (PCC) is commonly applied among the parameters during pattern construction. Zaki et al. [32] applied HMM to select the appropriate prediction model from several models they developed applying the adaptive neurofuzzy inference system (ANFIS). They obtained optimal state transition by four processing steps: initialization, recursion, termination, and backtracking. The last step analysed the previous step to determine the probability of the current state by using the Viterbi algorithm. Based on the log-likelihood of the initial model parameter, defined by expectation maximization (EM) algorithm, of HMM with the traffic pattern, a suitable congestion model was selected for prediction. Mishra et al. [23] applied the discretised multiple symbol HMM (MS-HMM) prediction model named future state prediction (FSP). They evaluated model adaptability for different road segments. A label was generated containing hidden states of MS-HMM, and the output was used for FSP to result in the next hidden state label.

In traffic engineering, especially while utilising probe vehicle data, HMM is very useful in map-matching. Sun et al. [45] applied HMM for mapping the trajectory of observed GPS points in nearby roads. These candidate points were taken as hidden states of HMM. The candidate points closer to the observation point had higher observation probability.

Transition probability of two adjacent candidates was also considered to avoid the misleading results generated from abrupt traffic situations.

HMM shows accuracy in selecting a traffic pattern or a traffic point. It has the advantage that it can deal with the data with outliers. However, points with a short sampling interval seem to be matched well, and long intervals and higher similar probe data decreased the model accuracy. Studies have found a significant mismatch for long sampling interval dataset and similar road networks.

The GPS tracking system has been widely developed in this era of the satellite. Thus, making HMM modelling is currently more relevant for map matching. Other sectors of transport where HMM is applied include traffic prediction [46], modified HMM for speed prediction [47], and traffic flow state transition [48], etc.

4.1.3. Gaussian Distribution.

Gaussian processes have proven to be a successful tool for regression problems. Formally, a Gaussian process is a collection of random variables, any finite number of which obeys a joint Gaussian prior distribution. For regression, the function to be estimated is assumed to be generated by an infinite-dimensional Gaussian distribution, and the observed outputs are contaminated by additive Gaussian noise.

Yang [29] applied Gaussian distribution for traffic congestion prediction in their study. This study was divided into three parts. First, the sensor ranking was done according to the volume quality by applying p test. In the second part of the study, the congestion occurring probability was determined from a statistics-based method. In the learning phase of this part, two Gaussian probability models were developed from two datasets for every point of interest. In the decision phase, on which model the input traffic volume value fitted was evaluated, and a prediction score presenting congestion state was determined from the ratio of two models. Finally, the probability of congestion occurring at the point of interest was found by combining and sorting the prediction score from all the ranked sensors. Zhu et al. [49] also presented the probability of traffic state distribution. Selection of mean and variance parameters of Gaussian distribution is an important step. In this study, the EM algorithm was applied for this purpose. The first step generated the log-likelihood expectation for the parameters, whereas the last step maximised it. Sun et al. [45] approximated the error in GPS location in the road with Gaussian Distribution, taking mean 0. The error was calculated from the actual GPS point, matching point on the road section, and standard deviation of GPS measurement error.

From the abovementioned studies, it is seen that the Gaussian distribution model has a useful application in reducing feature numbers without compromising the quality of the prediction results or for location error estimation while using GPS data. Gaussian distribution is also applied in traffic volume prediction [50], traffic safety [51], and traffic speed distribution variability [52].

4.1.4. Bayesian Network.

A Bayesian network (BN), also known as a causal model, is a directed graphical model for representing conditional independencies between a set of random variables. It is a combination of probability theory and graph theory and provides a natural tool for dealing with two problems that occur through applied mathematics and engineering—uncertainty and complexity [53].

Asencio-Cortés et al. [54] applied an ensemble of seven machine learning algorithms to compute the traffic congestion prediction. This methodology was developed as a binary classification problem applying the HIOCC algorithm. Machine learning algorithms applied in this study were K-nearest neighbour (K-NN), C4.5 decision trees (C4.5), artificial neural network (ANN) of backpropagation technique, stochastic gradient descent optimisation (SGD), fuzzy unordered rule induction algorithm (FURIA), Bayesian network (BN), and support vector machine (SVM). Three of these algorithms (C4.5, FURIA, and BN) can produce interpretable models of viewable knowledge. A set of ensembled learning algorithms were applied to improve the results found from these prediction models. The ensemble algorithm group included bagging, boosting (Ada-Boost M1), stacking, and Probability Threshold Selector (PTS). The authors found a significant improvement in Precision for BN after applying ensemble algorithms. On the other hand, Kim and Wang [34] applied BN to determine the factors that affect congestion initialization on different road sections. The developed model of this study gave a framework to assess different scenario ranking and prioritizing.

Bayesian network is seen to perform better with ensembled algorithms or while modified, e.g., other transport sectors of traffic flow prediction [55] and parameter estimation at signalised intersection [56, 57].

4.1.5. Others.

Other than the models mentioned above, the Kalman Filter (KF) is also a popular probabilistic algorithm. With the increment of available data, data fusion methods are becoming popular. The fusion of historical and real-time traffic data can achieve a higher level of traffic congestion prediction accuracy. In this regard, KF is commonly applied. Extended KF (EKF) is an extension of KF, which can be used to stochastically filter the nonlinear noises to improve the mean and covariance of an estimated state. Therefore, after data fusion, it updated the estimated covariance error by removing outliers [7].

Wen et al. [8] applied GA in traffic congestion prediction from spatiotemporal traffic environment. Temporal association rules were extracted from the traffic environment applying GA-based temporal association rules (GATARs). Their proposed Hybrid Temporal Association Rules Mining method (HTARM) included DBSCAN and GATAR methods. The DBSCAN application method was discussed previously in this article. While encoding using GATAR, road section number and congestion level were included in the chromosome. The decoding was done to obtain temporal association rules and was sorted according to confidence and support value in the rule pool. For both simulated and real-

world scenarios, the proposed HTARM method outperformed GATAR in terms of extracting temporal association rules and prediction accuracy. However, the cluster number difference showed a big difference in the two scenarios. Besides, with the increment of road network complexity, the prediction accuracy decreased.

Table 1 summarises the methodologies and different parameters used in various studies we have discussed so far.

4.2. Shallow Machine Learning.
Shallow machine learning (SML) algorithms include traditional and simple ML algorithms. These algorithms usually consist of a few, many times, one hidden layer. SML algorithms cannot extract features from the input, and features need to be defined beforehand. Model training can only be done after feature extraction. SML algorithms and their application in traffic congestion studies are discussed in this section and shown in Figure 6.

4.2.1. Artificial Neural Network.
Artificial neural network (ANN) was developed, mimicking the function of the human brain to solve different nonlinear problems. It is a first-order mathematical or computational model that consists of a set of interconnected processors or neurons. Figure 7 shows a simple ANN structure. Due to its easy implementation and efficient forecasting ability, ANN has become popular in the field of traffic congestion prediction research. Hopfield network, feedforward network, and backpropagation are the examples of ANN. Feedforward neural network (FNN) is the simplest NN, where the input data go to the hidden layer and from there to the output layer. Backpropagation neural network (BPNN) consists of feedforward and weight adjustment of the layers and is the most commonly applied ANN in transportation management. Xu et al. [31] applied BPNN to predict traffic flow, thus to evaluate congestion factor in their study. They proposed occupancy-based congestion factor (CRO) evaluation method with three other evaluated congestion factors based on mileage ratio of congestion (CMRC), road speed (CRS), and vehicle density (CVD). They also evaluated the effect of data-size on real-time rendering of road congestion. Complex road network with higher interconnections showed higher complication in simulation and rendering. The advantage of the proposed model was that it took little processing time for high sampling data rendering. The model can be used as a general congestion prediction model for different road networks. Some used hybrid NN for congestion prediction. Nadeem and Fowdur [11] predicted congestion in spatial space, applying the combination of one of six SML algorithms with NN. Six SML algorithms included moving average (MA), autoregressive integrated moving average (ARIMA), linear regression, second- and third-degree polynomial regression, and k-nearest neighbour (KNN). The model showing the least RMSE value was combined with BPNN to form hybrid NN. The hidden layer had seven neurons, which was determined by trial and error. However, it was a very preliminary level work. It did not show the effect of data increment in the accuracy.

Unlike the previous studies, those focused on traffic flow parameters to conduct traffic congestion prediction research; Ito and Kaneyasu [60] analysed drivers' behaviour in predicting congestion. They showed that vehicle operators act differently on different phases of the journey. They used one layered BPNN to learn the behaviour of female drivers and extract travel phase according to that. The results showed an average efficiency of 82% in distinguishing the travel phase.

ANN is a useful machine learning model which has a flexible structure. The neurons of the layer can be adapted according to the input data. As mentioned above, a general model can be developed and applied for different road types by using the advantage of nonlinearity capturing ability of ANN. However, ANN requires larger datasets than the probabilistic reasoning models, which results in high complexity.

ANN shows great potential in diverse parameter analysis. ANN is the only model that has recently been applied for driver behaviour analysis for traffic congestion. ANN is popular in every section of transport- traffic flow prediction [61, 62], congestion control [63], driver tiredness [64], and vehicle noise [65, 66].

4.2.2. Regression Model.
Regression is a statistical supervised ML algorithm. It models the prediction real numbered output value based on the independent input numerical variable. Regression models can be further divided according to the number of input variables. The simplest regression model is linear regression with one input feature. When the feature number increases, the multiple regression model is generated.

Jiwan et al. [27] developed a multiple linear regression analysis (MLRA) model using weather data and traffic congestion data after preprocessing using Hadoop. At first, a single regression model was developed for all the variables using R. After a 3-fold reduction process, only ten variables were determined to form the final MLRA model. Zhang and Qian [22] conducted an interesting approach to predict morning peak hour congestion using household electricity usage patterns. They used LASSO regression to correlate the pattern features using the advantage of linearly related critical feature selection capability.

On the other hand, Jain et al. [33] developed both linear and exponential regression model using IBM SPSS software to find the relevant variables. The authors converted heterogenous vehicles into passenger car unit (PCU) for simplification. Three independent variables were considered to estimation origin-destination- (O-D-) based congestion measures. They used PCC to evaluate the correlation among the parameters. However, simply averaging O-D node parameters may not provide the actual situation of dynamic traffic patterns.

Regression models consist of some hidden coefficients, which are determined in the training phase. The most applied regression model is the autoregressive integrated moving average (ARIMA). ARIMA has three parameters- p, d, and q. "p" is the auto regressive order that refers to how

TABLE 1: Traffic congestion prediction studies in probabilistic reasoning.

Methodology	Road type	Data source	Input parameters	Target domain	No. of congestion state levels*	Reference
Hierarchical fuzzy rule-based system			Occupancy	Speed	2	Zhang et al. [30]
	Highway corridor	Sensor	Speed	Speed		Lopez-garcia et al. [37]
Evolutionary fuzzy rule learning			Traffic flow	Traffic density		Onieva et al. [28]
Mamdani-type fuzzy logic inference	Highway, trunk road, branch road	—	Speed		4	Cao and Wang [3]
Fuzzy inference	Highway corridor	Camera	Density Travel time Traffic flow Speed	Congestion Index		Wang et al. [58]
Fuzzy comprehensive evaluation	Highway corridor	Probe	Traffic volume Speed	Saturation Density speed	5	Kong et al. [4] Yang et al. [5]
	Highway network	Sensor	Emission matrix	Traffic pattern selection	—	Zaki et al. [32]
Hidden Markov model			Emission matrix Transition matrix	Traffic pattern determination	—	Zaki et al. [25]
	Main road	Probe	Observation probability Transition probability	Mapping GPS data	—	Sun et al. [45]
Gaussian distribution	Highway corridor	Sensor	Traffic volume	Optimal feature selection	—	Yang [29]
	Build-up area	Simulation	Road and bus increment			Yi Liu et al. [59]
	Bridge	Sensor	Intensity Occupation Average speed Average distance Network direction	Congestion probability	—	Asencio-Cortés et al. [54]
Bayesian network			Day and time weather Incidents			
	Highway network	Sensor	Traffic flow Occupancy Speed Level of service Congestion state	Congestion probability		Kim and Wang [34]
Extended Kalman filter	Highway	Camera	Travel time	Data fusion	—	Adetiloye and Awasthi [7]

The table accumulates the data source, scope of the study area, input and resulting parameters, and how many cognitive traffic states were considered in the studies.*2 = free/congested, 4 = free/light/medium/severe, 5 = very free/free/light/medium/severe

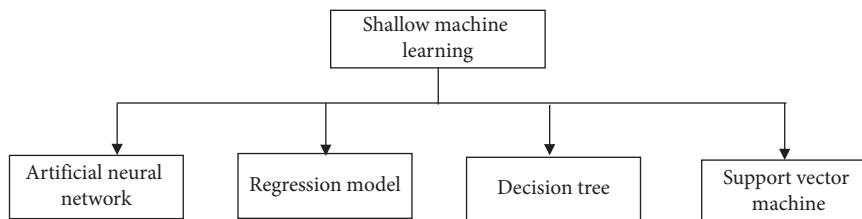FIGURE 6: Subdivision of shallow machine learning models.

many lags of the independent variable needs to be considered for prediction. Moving average order "q" presents the lag prediction error numbers. Lastly, "d" is used to make the time-series stationary. Alghamdi et al. [67] took d as 1 as one differencing order could make the model stationary. Next, they applied the autocorrelation function (ACF) and
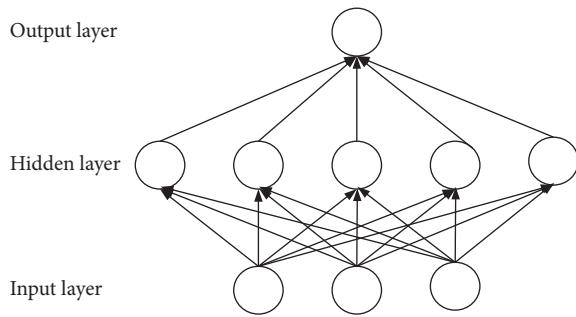
FIGURE 7: A simple ANN structure.

the partial autocorrelation function (PACF) along with the minimum information criteria matrix to determine the values of p and q. They only took the time dimension into account. However, the results inclined with the true pattern for only one week and needed to be fine-tuned considering prediction errors. Besides, the study did not consider the spatial dimension.

Regression models are useful to be applied for time series problems. Therefore, regression models are suitable for traffic forecasting problems. However, these models are not reliable for nonlinear, rapidly changing the multidimension dataset. The results need to be modified according to prediction errors.

However, as already and further will be discussed in this article, most of the studies used different regression models to validate their proposed model [6, 11, 25, 68, 69].

With the increment of dataset and complexity associated with it, regression models are becoming less popular in traffic congestion prediction. Currently, regression models are frequently used by modifying with other machine learning algorithms, e.g., ANN and kernel functions. Some other sectors' regression models are applied including hybrid ARIMA in traffic speed prediction for specific vehicle type (Wang et al. [70], traffic volume prediction [71], and flow prediction applying modified ARIMA [72].

4.2.3. Decision Tree.
A decision tree is a model that predicts an output based on several input variables. There are two types of trees: the classification tree and the regression tree. When these two trees merge, a new tree named classification and regression tree (CART) generates. Decision tree uses the features extracted from the entire dataset. Random forest is a supervised ML classification algorithm that is the average of multiple decision tree results. The features are randomly used while developing decision trees. It uses a vast amount of CART decision trees. The decision trees vote for the predicted class in a random forest model.

Wang et al. [9] proposed a probabilistic method of exploiting information theory tools of entropy and Fano's inequality to predict road traffic pattern and its associated congestion for urban road segments with no prior knowledge on the O-D of the vehicle. They incorporated road congestion level into time series for mapping the vehicle state into the traffic conditions. As interval influenced the

predictability, an optimal segment length and velocity was found. However, with less available data, an increased number of segments increased the predictability. Another traffic parameter, travel time, was used to find CI by Liu and Wu [73]. They applied the random forest ML algorithm to forecast traffic congestion states. At first, they extracted 100 sample sets to construct 100 decision trees by using bootstrap. The number of feature attributes was determined as the square root of the total number of features. Chen et al. [16] also applied the CART method for prediction and classification of traffic congestion. The authors applied Moran's I method to analyse the spatiotemporal correlation among different road network traffic flow. The model showed effectiveness compared with SVM and K-means algorithm.

Decision tree is a simple classification problem-solving model that can be applied for multifeature data, e.g., Liu and Wu [73] applied weather condition, road condition, time period, and holiday as the input variables. This model's knowledge can be represented in the form of IF-THEN rules, making it an easily interpretable problem. It is also needed to be kept in mind that the classification results are usually binary and therefore, not suitable where the congestion level is required to be known. Other sectors of transport, where decision tree models applied are traffic prediction [74] and traffic signal optimisation with Fuzzy logic [75].

4.2.4. Support Vector Machine.
The support vector machine (SVM) is a statistical machine learning method. The main idea of this model is to map the nonlinear data to a higher dimensional linear space where data can be linearly classified by hyperplane [1]. Therefore, it can be very useful in traffic flow pattern identification for traffic congestion prediction. Tseng et al. [13] determined travel speed in predicting real-time congestion applying SVM. They used Apache Storm to process big data using spouts and bolts. Traffic, weather sensors, and events collected from social media of close proximity were evaluated together by the system. They categorised vehicle speed into classes and referred them as labels. Speed of the previous three intervals was used to train the proposed model. However, the congestion level categorised from 0 to 100 does not carry a specific knowledge of the severity of the level, especially to the road users. Increment in training data raised accuracy and computational time. This may ultimately make it difficult to make real-time congestion prediction.

Traffic flow shows different patterns based on the traffic mixture or time of the day. SVM is applied to identify the appropriate pattern. Currently modified SVM mostly has its application in other sectors as well, e.g., freeway exiting traffic volume prediction [58], traffic flow prediction [76], and sustainable development of transportation and ecology [77].

Most of the studies compared their developed model with SVM [22, 78, 79]. Deep machine learning (DML) algorithms showed better results compared to SVM. Table 2 refers to the studies under this section.

TABLE 2: Traffic congestion prediction studies in shallow machine learning.

Methodology	Road type	Data source	Input parameters	Target domain	No. of congestion state levels	Reference
Artificial neural network	Road network	Sensor Simulation	Occupancy Density	Congestion factor	3	Xu et al. [31]
	Highway corridor	Sensor	Distance	Speed	2	Nadeem and Fowdur [11]
		Simulation	Speed Speed Throttle opening Steering input angle	Traffic congestion state	2	Ito and Kaneyasu [60]
Regression model	Highway corridor	Sensor	Temperature Humidity Rainfall Traffic speed Time	Traffic congestion score	—	Jiwan et al. [27]
	Arterial road Subarterial road	Camera	Volume Speed	Congestion Index	4	Jain et al. [33]
Decision tree	Ring road	Probe	Average speed	Traffic predictability	—	Wang et al. [9]
	Road network		Speed Trajectory	Moran index	5	Chen et al. [16]
Support vector machine	Highway corridor	Sensor	Speed Density Traffic volume difference Rainfall volume	Travel speed	—	Tseng et al. [13]

4.3. Deep Machine Learning. DML algorithms consist of several hidden layers to process nonlinear problems. The most significant advantage of these algorithms is they can extract features from the input data without any prior knowledge. Unlike SML, feature extraction and model training are done together in these algorithms. DML can convert the vast continuous and complex traffic data with limited collection time horizon into patterns or feature vectors. From last few years, DML has become popular in traffic congestion prediction studies. Traffic congestion studies that used DML algorithms are shown in Figure 8 and discussed in this section.

4.3.1. Convolutional Neural Network. Convolutional neural network (CNN) is a commonly applied DML algorithm in traffic engineering. Due to the excellent performance of CNN in image processing, while applying in traffic prediction, traffic flow data are converted into a 2-D matrix to process. There are five main parts of a CNN structure in transportation: the input layer, convolution layer, pool layer, full connection layer, and output layer. Both the convolution and pooling layer extracts important features. The depth of these two layers differs in different studies. Majority of the studies converted traffic flow data into an image of a 2-D matrix. In the studies performed by Ma et al. [80] and Sun et al. [45], each component of the matrix represented average traffic speed on a specific part of the time. While tuning CNN parameters, they selected a convolutional filter size of (3×3) and max-pooling of size (2×2) of 3 layers according to parameter settings of LeNet and AlexNet and loss of information measurement.

Whereas Chen et al. [68] used a five-layered convolution of filter size of (2×2) without the pooling layer. The authors applied a novel method called convolution-based deep neural network modelling periodic traffic data (PCNN). The study folded the time-series to generate the input combining real-time and historical traffic data. To capture the correlation of a new time slot with the immediate past, they duplicated the congestion level of the last slot in the matrix. Zhu et al. [49] also applied five convolution-pooling layers as well as (3×3) and (2×2) sizes, respectively. Along with temporal and spatial data, the authors also incorporated time interval data to produce a 3-D input matrix. Unlike these studies, Zhang et al. [6] preprocessed the raw data by performing a spatiotemporal cross-correlation analysis of traffic flow sequence data using PCC. Then, they applied a model named spatiotemporal feature selection algorithm (STFSA) on the traffic flow sequence data to select the feature subsets as the input matrix. A 2-layered CNN with the convolutional and pooling size as same as the previous studies was used. However, STFSA considers its heuristics, biases, and trade-offs and does not guarantee optimality.

CNN shows good performance, where a large dataset is available. It has excellent feature learning capability with less time-consuming classification ability. Therefore, CNN can be applied where the available dataset can be converted into an image. CNN is applied in traffic speed prediction [81], traffic flow prediction [6], and modified CNN with LSTM is also applied for traffic prediction [82]. However, as mentioned above, no model depth and parameter selection strategies are available.
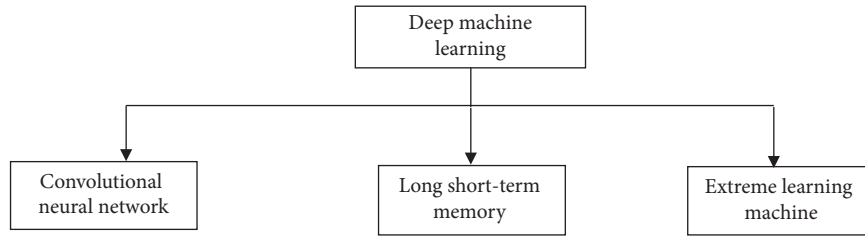
FIGURE 8: Subdivision of deep machine learning models.

4.3.2. Recurrent Neural Network.

Recurrent neural network (RNN) has a wide usage in the sequential traffic data processing by considering the influence of the related neighbour (Figure 9). Long short-term memory (LSTM) is a branch of RNN. In the hidden layer of LSTM, there is a memory block that includes four NN layers, which stores and regulates the information flow. In recent years, with different data collection systems with extended intervals, LSTM has become popular. Due to this advantage, Zhao et al. [12] developed an LSTM model consisted of three hidden layers and ten neurons using long interval data. They set an adequate target and fine-tuned the parameters until the training model stabilized. The authors also applied the congestion index and classification (CI-C) model to classify the congestion by calculating CI from LSTM output data. Most of the studies use the equal interval of CI to divide congestion states. This study did two more intervals of natural breakpoint and geometric interval to find that the latest provided the most information from information entropy. Lee et al. [69] applied 4 layers with 100 neurons LSTM model of 3D matrix input. The input matrix element contained a normalised speed to shorten the training time. While eliminating the dependency, the authors found that a random distribution of target road speed and more than optimally connected roads in the matrix reduced the performance. To eliminate the limitation of temporal dependency, Yuan-Yuan et al. [79] trained their model in the batch learning approach. The instance found from classifying test dataset was used to train the model in an online framework. Some studies introduced new layers to modify the LSTM model for feature extraction. Zhang et al. [83] introduced an attention mechanism layer between LSTM and prediction layer that enabled the feature extraction from a traffic flow data sequence and captured the importance of a traffic state. Di et al. [84] introduced convolution that provides an input to the LSTM model to form the CPM-ConvLSTM model. All the studies applied the one-hot method to convert the input parameters. Adam, stochastic gradient descent (SGD), and leakage integral echo state network (LiESN) are a few optimisation methods applied to fine-tune the outcome.

A few studies combined RNN with other algorithms while dealing with vast parameters of the road network. In this regard, Ma et al. [85] applied the RNN and restricted Boltzmann machine (RNN-RBM) model for networkwide spatiotemporal congestion prediction. Here, they used conditional RBM to construct the proposed deep architecture, which is designed to process the temporal sequence by providing a feedback loop between visible layer and hidden
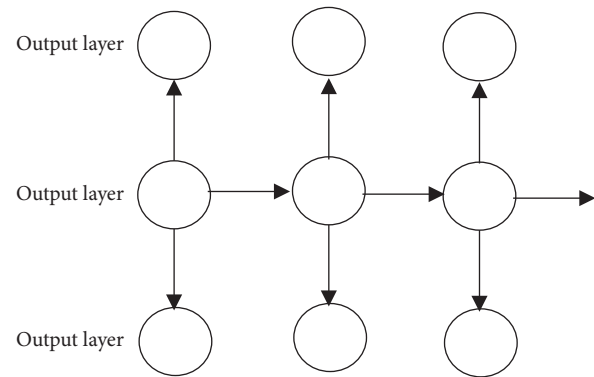

FIGURE 9: A simple RNN structure.

layer. The congestion state was analysed from traffic speed and was represented in binary format in a matrix as input. Also, Sun et al. [45] combined RNN of three hidden layers, with its two other variants: LSTM and gated recurrent unit (GRU). The hidden layers included the memory block characteristic of LSTM, and the cell state and hidden state were incorporated by GRU.

As the sample size is increasing vastly, RNN is becoming popular as a current way of modelling. RNN has a short-term memory. This characteristic of RNN helps to model nonlinear time series data. The training of RNN is also straight forward, similar to multilayer FNN. However, this training may become difficult due to the conversion in a deep architecture with multiple layers in long-term dependency. In case of long-term dependency problems, LSTM is becoming more suitable to be applied as LSTM can remember information for a long period of time. RNN has its application in other sectors of the transport too, e.g., traffic passenger flow prediction [86], modified LSTM in real time crash prediction [87], and road-network traffic prediction [88].

4.3.3. Extreme Learning Machine.

In recent years, a novel learning algorithm called the extreme learning machine (ELM) is proposed for training the single layer feed-forward neural network (SLFN). In ELM, input weights and hidden biases are assigned randomly instead of being exhaustively tuned. Therefore, ELM training is fast. Therefore, taking this advantage into account, Ban et al. [19] applied the ELM model for real-time traffic congestion prediction. They determine CI using the average travel speed. A 4-fold cross-validation was done to avoid noise in raw data. The model

TABLE 3: Traffic congestion prediction studies in deep machine learning.

Methodology	Road type	Data source	Input parameters	Target domain	No. of congestion state levels	Reference
Convolutional neural networks	Road network	Probe	Average traffic speed	Speed	3	Ma et al. [80]
				Average traffic speed	5	Sun et al. [45]
		Camera	Congestion level	Congestion level	3	Chen et al. [68]
	Highway corridor	Sensor	Traffic flow	Traffic flow	—	Zhang et al. [93]
	Road section	Probe	Weather data Congestion time	Congestion time	5	Zhao et al. [12]
	Arterial road	Online	Congestion level	Congestion level		Yuan-Yuan et al. [79]
Recurrent neural network	Road network	Camera	Spatial similarity feature	Speed	3	Lee et al. [69]
		Sensor	Speed			
		Survey	Peak hour Speed			
	Highway corridor	Sensor	Travel time Volume	Congestion level	4	Zhang et al. [83]
			Congestion state Current time Road traffic state cluster	Congestion state	2	Ma et al. [85]
Extreme learning machine	Road network	Probe	Last congestion index Road type Number of adjacent roads	Congestion Index	—	Ban et al. [19]

found optimal hidden nodes to be 200 in terms of computational cost in the study. An extension of this study was done by Shen et al. [78] and Shen et al. [89] by applying a kernel-based semisupervised extreme learning machine (kernel-SSELM) model. This model can deal with the unlabelled data problem of ELM and the heterogenous data influence. The model integrated small-scaled labelled data of transportation personnel and large-scaled unlabelled traffic data to evaluate urban traffic congestion. ELM speeded up the processing time, where kernel function optimized the accuracy and robustness of the whole model. However, real-time labelled data collection was quite costly in terms of human resources and working time, and the number of experts for congestion state evaluation should have been more. Another modification of EML was applied by Yiming et al. [20]. They applied asymmetric extreme learning machine cluster (S-ELM-cluster) model for short-term traffic congestion prediction by determining the CI. The authors divided the study area and implemented submodels processing simultaneously for fast speed.

The ELM model has the advantage in processing large scale data learning at high speed. ELM works better with labelled data. Where both labelled and unlabelled data are available, semisupervised ELM has shown good prediction accuracy, as it was seen from the studies. Other sectors where ELM was applied included air traffic flow prediction [90], traffic flow prediction [91], and traffic volume interval prediction [92].

Other than the models already discussed, Zhang et al. [93] proposed a deep autoencoder-based neural network model with symmetry of four layers for the encoder and the decoder to learn temporal correlations of a transportation network. The first component encoded the vector representation of historical congestion levels and their correlation. They then decoded to build a representation of congestion levels for the future. The second component of DCPS used two dense layers; those converted the output from the decoder to calculate a vector representation of congestion level. However, the process lost information as the congestion level of all the pixels was averaged. This approach needed high iteration and was computationally expensive as all the pixels regardless of roads were considered. Another study applied a generalised version of recurrent neural network named recursive neural network. The difference between these two is, in recurrent NN, weights are shared along the data sequence. Whereas recursive NN is a single neuron model; therefore, weights are shared at every node. Huang et al. [94] applied a recursive NN algorithm named echo state network (ESN). This model consists of an input layer, reservoir network, and output layer. The reservoir layer constructs the rules that connected prediction origin and forecasting horizon. As the study took a large study area with vast link number, they simplified the training rule complexity applying recursive NN. Table 3 summarises some studies.

5. Discussion and Research Gaps

Research in traffic congestion prediction is increasing exponentially. Among the two sources, most of the studies used stationary sensor/camera data. Although sensor data cannot capture the dynamic traffic change, frequent change

in source makes it complicated to evaluate the flow patterns for probe data [95]. Data collection horizon is an important factor in traffic congestion studies. The small horizon of a few days [3] cannot capture the actual situation of the congestion as traffic is dynamic. Other studies that used data for a few months showed the limitation of seasonality [22, 67].

The condition of the surrounding plays an important factor in traffic congestion. A few studies focused on these factors. Two studies considered social media contribution in input parameter [7, 13], and five considered weather condition [12, 13, 27, 34, 73]. Events, e.g., national event, school holiday, and popular sports events, play a big role in traffic congestion. For example, Melbourne, Australia, has two public holidays before and during two most popular sports events of the country. The authorities close a few traffic routes to tackle the traffic and the parade, resulting in traffic congestion. Therefore, more focus must be put in including these factors while forecasting.

Dealing with missing data is a challenge in the data processing. Some excluded the respective data altogether [29], others applied different methods to retrieve the data [59, 85], and some replaced with other data [45]. Missing data imputation can be a useful research scope in transportation engineering.

Machine learning algorithm, especially DML models, is developed with time. This shows a clear impact on the rise of their implementation in traffic congestion forecasting (Figure 10).

Probabilistic reasoning algorithms were mostly applied for a part of the prediction model, e.g., map matching and optimal feature number selection. Fuzzy logic is the most widely used algorithm in this class of algorithms. From other branches, ANN and RNN are the mostly applied models. Most of the studies that applied hybrid or ensembled models belong to probabilistic and shallow learning class. Only two studies applied hybrid deep learning models while predicting networkwide congestion. Tables 4, 5–6 summarize the advantage and weaknesses of the algorithms of different branches.

Among all DML models, RNN is more suitable for time series prediction. In a few studies, RNN performed better than CNN as the gap between the traffic speeds in different classes was very small [12, 69]. However, due to little research in traffic congestion field, a lot of new ML algorithms are yet to be applied.

SML models showed better results than DML while forecasting traffic congestion in the short-term, as SML can process linearity efficiently and linear features have more contribution to traffic flow in short-term. All the short-term forecasting studies discussed in this article applying SML showed promising results. At the same time, DML models showed good accuracy as these models can handle both
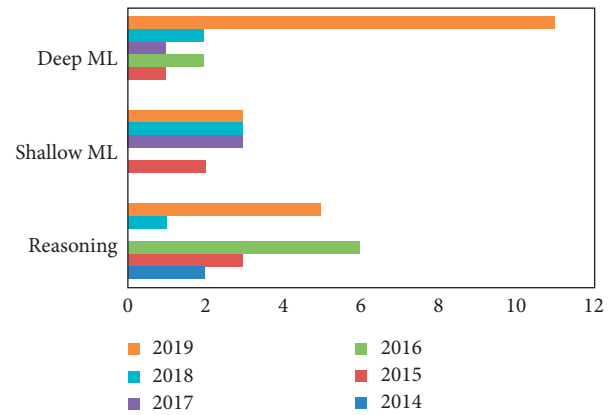

FIGURE 10: Application of AI models with time.

linear and nonlinear features efficiently. Besides, real-time congestion prediction cannot afford high computation time. Therefore, models taking a short computational time are more effective in this case.

6. Future Direction

Traffic congestion is a promising area of research. Therefore, there are multiple directions to conduct in future research.

Numerous forecasting models have already been applied in road traffic congestion forecasting. However, with the newly developed forecasting models, there is more scope to make the congestion prediction more precise. Also, in this era of information, the use of increased available traffic data by applying the newly developed forecasting models can improve the prediction accuracy.

The semisupervised model was applied only for the EML model. Other machine learning algorithms should be explored for using both labelled and unlabelled data for higher prediction accuracy. Also, a limited number of studies have focused on real-time congestion forecasting. In future, researches should pay attention to real-time traffic congestion estimation problem.

Another future direction can be focusing on the level of traffic congestion. A few studies have divided the traffic congestion into a few states. However, for better traffic management, knowing the grade of congestion is essential. Therefore, future researches should focus on this. Besides, most studies focused on only one traffic parameter to forecast congestion for congestion prediction. This can be an excellent future direction to give attention to more than one parameter and combining the results during congestion forecasting to make the forecasting more reliable.

TABLE 4: The strength and weakness of the models of probabilistic reasoning.

Methodology	Advantages	Disadvantages
Fuzzy logic	(i) It converts the binary value into the linguistic description hence portraying the traffic congestion state.	(i) No appropriate membership function shape selection method exists.
	(ii) iIt can portray more than two states.	(ii) Traffic pattern recognition capability is not as durable as ML algorithms.
	(iii) As it does not need an exact crisp input, it can deal with uncertainty.	(iii) Traffic state may not match the actual traffic state as the outcome is not exact.
Hidden Markov model	(i) The model can overcome noisy measurements.	(i) Accuracy decreases with scarce temporal probe trajectory data
	(ii) Can efficiently learn from non-preprocessed data. (iii) Can evaluate multiple hypotheses of the actual mapping simultaneously.	(ii) Not suitable in case of missing dataset.
Gaussian mixture model	(i) Can do traffic parameter distribution over a period as a mixture regardless of the traffic state. (ii) Can overcome the limitation of not being able to account for multimodal output by a single Gaussian process.	(i) Optimization algorithm used with GMM must be chosen cautiously. (ii) Results may show wrong traffic patterns due to local optima limitation and lack of traffic congestion threshold knowledge of the optimisation algorithm.
Bayesian network	(i) It can understand the underlying relationship between random variables.	(i) Computationally expensive.
	(ii) It can model and analyse traffic parameters between adjacent road links.	(ii) The model performs poorly with the increment in data.
	(iii) The model can work with incomplete data.	(iii) The model represents one-directional relation between variables only.

TABLE 5: The strength and weakness of the models of shallow machine Learning.

Methodology	Advantages	Disadvantages
Artificial neural network	(i) It is an adaptive system that can change structure based on inputs during the learning stage [96].	(i) BPNN requires vast data for training the model due to the parameter complexity resulting from its parameter nonsharing technique [97].
	(ii) It features defined early, FNN shows excellent efficiency in capturing the nonlinear relationship of data.	(ii) The training convergence rate of the model is slow.
Regression model	(i) Models are suitable for time series problems.	(i) Linear models cannot address nonlinearity, making it harder to solve complex prediction problems.
	(ii) Traffic congestion forecasting problems can be easily solved.	(ii) Linear models are sensitive to outliers.
	(iii) ARIMA can increase accuracy by maintaining minimum parameters.	(iii) Computationally expensive.
	(iv) Minimum complexity in the model.	(iv) ARIMA cannot deal multifeature dataset efficiently. (v) ARIMA cannot capture the rapidly changing traffic flow [8].
Support vector machine	(i) It is efficient in pattern recognition and classification.	(i) The improperly chosen kernel function may result in an inaccurate outcome.
	(ii) A universal learning algorithm that can diminish the classification error probability by reducing the structural risk [1].	(ii) Unstable traffic flow requires improved prediction accuracy of SVM.
	(iii) It does not need a vast sample size.	(iii) It takes high computational time and memory.

TABLE 6: The strength and weakness of the models of deep machine learning.

Methodology	Advantages	Disadvantages
Convolutional neural networks	(i) Capable of learning features from local connections and composing them into high-level representation. (ii) Classification is less time-consuming. (iii) Can automatically extract features.	(i) Computationally expensive as a huge kernel is needed for feature extraction. (ii) A vast dataset is required. (iii) Traffic data needs to be converted to an image. (iv) No available strategies are available on CNN model depth and parameter selection.
Recurrent neural network	(i) Shows excellent performance in processing sequential data flow. (ii) Efficient in sequence classification. (iii) Efficient in processing time-series with long intervals and postponements.	(i) Long-term dependency results in bad performance. (ii) No available firm guideline in dependency elimination.
Extreme learning machine	(i) Fast learning speed (ii) Can avoid local minima. (iii) Modified models are available to deal with an unlabelled data problem.	(i) Training time increases with the hidden node rise. (ii) Unlabelled data problem. (iii) May produce less accurate results.

7. Conclusions

Traffic congestion prediction is getting more attention from the last few decades. With the development of infrastructure, every country is facing traffic congestion problem. Therefore, forecasting the congestion can allow authorities to make plans and take necessary actions to avoid it. The development of artificial intelligence and the availability of big data have led researchers to apply different models in this field. This article divided the methodologies in three classed. Although probabilistic models are simple in general, they become complex while different factors that affect traffic congestion, e.g., weather, social media, and event, are considered. Machine learning, especially deep learning, has the benefit in this case. Therefore, deep learning algorithms became more popular with time as they can assess a large dataset. However, a wide range of machine learning algorithms are yet to be applied. Therefore, a vast opportunity of research in the field of traffic congestion prediction still prevails.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] Z. Shi, *Advanced Artificial Intelligence*, World Scientific, Singapore, 2011.

[2] M. S. Ali, M. Adnan, S. M. Noman, and S. F. A. Baqueri, "Estimation of traffic congestion cost-A case study of a major arterial in karachi," *Procedia Engineering*, vol. 77, pp. 37–44, 2014.

[3] W. Cao and J. Wang, "Research on traffic flow congestion based on Mamdani fuzzy system," *AIP Conference Proceedings*, vol. 2073, 2019.

[4] X. Kong, Z. Xu, G. Shen, J. Wang, Q. Yang, and B. Zhang, "Urban traffic congestion estimation and prediction based on floating car trajectory data," *Future Generation Computer Systems*, vol. 61, pp. 97–107, 2016.

[5] Q. Yang, J. Wang, X. Song, X. Kong, Z. Xu, and B. Zhang, "Urban traffic congestion prediction using floating car trajectory data," in *Proceedings of the International Conference on Algorithms and Architectures for Parallel Processing*, pp. 18–30, Springer, Zhangjiajie, China, November 2015.

[6] W. Zhang, Y. Yu, Y. Qi, F. Shu, and Y. Wang, "Short-term traffic flow prediction based on spatio-temporal analysis and CNN deep learning," *Transportmetrica A: Transport Science*, vol. 15, no. 2, pp. 1688–1711, 2019.

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

[8] F. Wen, G. Zhang, L. Sun, X. Wang, and X. Xu, "A hybrid temporal association rules mining method for traffic congestion prediction," *Computers & Industrial Engineering*, vol. 130, pp. 779–787, 2019.

[9] J. Wang, Y. Mao, J. Li, Z. Xiong, and W.-X. Wang, "Predictability of road traffic and Congestion in urban areas," *PLoS One*, vol. 10, no. 4, Article ID e0121825, 2015.

[10] Z. He, G. Qi, L. Lu, and Y. Chen, "Network-wide identification of turn-level intersection congestion using only low-frequency probe vehicle data," *Transportation Research Part C: Emerging Technologies*, vol. 108, pp. 320–339, 2019.

[11] K. M. Nadeem and T. P. Fowdur, "Performance analysis of a real-time adaptive prediction algorithm for traffic congestion," *Journal of Information and Communication Technology*, vol. 17, no. 3, pp. 493–511, 2018.

[12] H. Zhao, X. Jizhe, L. Fan, L. Zhen, and L. Qingquan, "A peak traffic Congestion prediction method based on bus driving time," *Entropy*, vol. 21, no. 7, p. 709, 2019.

[13] F.-H. Tseng, J.-H. Hsueh, C.-W. Tseng, Y.-T. Yang, H.-C. Chao, and L.-D. Chou, "Congestion prediction with big data for real-time highway traffic," *IEEE Access*, vol. 6, pp. 57311–57323, 2018.

[14] D. Li, J. Deogun, W. Spaulding, and B. Shuart, "Towards missing data imputation: a study of fuzzy K-means Clustering method," in *Rough Sets and Current Trends in Computing*, S. Tsumoto, R. Słowiński, J. Komorowski, and J. W. Grzymała-Busse, Eds., pp. 573–579, Springer, Berlin, Germany, 2004.

[15] Y. Yang, Z. Cui, J. Wu, G. Zhang, and X. Xian, "Fuzzy c-means clustering and opposition-based reinforcement learning for traffic congestion identification," *Journal of Information & Computational Science*, vol. 9, no. 9, pp. 2441–2450, 2012.

[16] Z. Chen, Y. Jiang, D. Sun, and X. Liu, "Discrimination and prediction of traffic congestion states of urban road network based on spatio-temporal correlation," *IEEE Access*, vol. 8, pp. 3330–3342, 2020.

[17] Y. Guo, L. Yang, S. Hao, and J. Gao, "Dynamic identification of urban traffic congestion warning communities in heterogeneous networks," *Physica A: Statistical Mechanics and Its Applications*, vol. 522, pp. 98–111, 2019.

[18] A. Elfar, A. Talebpour, and H. S. Mahmassani, "Machine learning approach to short-term traffic congestion prediction in a connected environment," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2672, no. 45, pp. 185–195, 2018.

[19] X. Ban, C. Guo, and G. Li, "Application of extreme learning machine on large scale traffic congestion prediction," *Proceedings of ELM-2015*, vol. 1, pp. 293–305, 2016.

[20] X. Yiming, B. Xiaojuan, L. Xu, and S. Qing, "Large-scale traffic Congestion prediction based on the symmetric extreme learning machine Cluster fast learning method," *Symmetry*, vol. 11, no. 6, p. 730, 2019.

[21] Y. Zheng, Y. Li, C.-M. Own, Z. Meng, and M. Gao, "Real-time predication and navigation on traffic congestion model with equilibrium Markov chain," *International Journal of Distributed Sensor Networks*, vol. 14, no. 4, Article ID 155014771876978, 2018.

[22] P. Zhang and Z. Qian, "User-centric interdependent urban systems: using time-of-day electricity usage data to predict morning roadway congestion," *Transportation Research Part C: Emerging Technologies*, vol. 92, pp. 392–411, 2018.

[23] P. Mishra, R. Hadfi, and T. Ito, "Adaptive model for traffic congestion prediction," in *Proceedings of the International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*, vol. 9799, pp. 782–793, Morioka, Japan, 2016.

[24] F. Li, J. Gong, Y. Liang, and J. Zhou, "Real-time congestion prediction for urban arterials using adaptive data-driven methods," *Multimedia Tools and Applications*, vol. 75, no. 24, pp. 17573–17592, 2016.

[25] J. F. Zaki, A. Ali-Eldin, S. E. Hussein, S. F. Saraya, and F. F. Areed, "Traffic congestion prediction based on Hidden Markov Models and contrast measure," *Ain Shams Engineering Journal*, vol. 11, no. 3, p. 535, 2020.

[26] P. Jiang, L. Liu, L. Cui, H. Li, and Y. Shi, "Congestion prediction of urban traffic employing SRBDP," in *Proceedings of the 2017 IEEE International Symposium on Parallel and Distributed Processing with Applications and 2017 IEEE International Conference on Ubiquitous Computing and Communications (ISPA/IUCC)*, pp. 1099–1106, IEEE, Guangzhou, China, 2017.

[27] L. Jiwan, H. Bonghee, L. Kyungmin, and J. Yang-Ja, "A prediction model of traffic congestion using weather data," in *Proceedings of the 2015 IEEE International Conference on Data Science and Data Intensive Systems*, pp. 81–88, Sydney, NSW, Australia, December 2015.

[28] E. Onieva, P. Lopez-Garcia, A. D. Masegosa, E. Osaba, and A. Perallos, "A comparative study on the performance of evolutionary fuzzy and Crisp rule based Classification methods in Congestion prediction," *Transportation Research Procedia*, vol. 14, pp. 4458–4467, 2016.

[29] S. Yang, "On feature selection for traffic congestion prediction," *Transportation Research Part C: Emerging Technologies*, vol. 26, pp. 160–169, 2013.

[30] X. Zhang, E. Onieva, A. Perallos, E. Osaba, and V. C. S. Lee, "Hierarchical fuzzy rule-based system optimized with genetic algorithms for short term traffic congestion prediction," *Transportation Research Part C: Emerging Technologies*, vol. 43, pp. 127–142, 2014.

[31] Y. Xu, L. Shixin, G. Keyan, Q. Tingting, and C. Xiaoya, "Application of data science technologies in intelligent prediction of traffic Congestion," *Journal of Advanced Transportation*, 2019.

[32] J. Zaki, A. Ali-Eldin, S. Hussein, S. Saraya, and F. Areed, "Time aware hybrid hidden Markov models for traffic Congestion prediction," *International Journal on Electrical Engineering and Informatics*, vol. 11, no. 1, pp. 1–17, 2019.

[33] S. Jain, S. S. Jain, and G. Jain, "Traffic congestion modelling based on origin and destination," *Procedia Engineering*, vol. 187, pp. 442–450, 2017.

[34] J. Kim and G. Wang, "Diagnosis and prediction of traffic Congestion on urban road networks using bayesian networks," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2595, no. 1, pp. 108–118, 2016.

[35] W.-X. Wang, R.-J. Guo, and J. Yu, "Research on road traffic congestion index based on comprehensive parameters: taking Dalian city as an example," *Advances in Mechanical Engineering*, vol. 10, no. 6, Article ID 168781401878148, 2018.

[36] E. Onieva, V. Milanes, J. Villagra, J. Perez, and J. Godoy, "Genetic optimization of a vehicle fuzzy decision system for intersections," *Expert Systems with Applications*, vol. 39, no. 18, pp. 13148–13157, 2012.

[37] P. Lopez-Garcia, E. Onieva, E. Osaba, A. D. Masegosa, and A. Perallos, "A hybrid method for short-term traffic Congestion forecasting using genetic algorithms and Cross entropy," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 2, pp. 557–569, 2016.

[38] A. Daissaoui, A. Boulmakoul, and Z. Habbas, "First specifications of urban traffic-congestion forecasting models," in *Proceedings of the 27th International Conference on Microelectronics (ICM 2015).*, pp. 249–252, Casablanca, Morocco, December 2015.

[39] M. Collotta, L. Lo Bello, and G. Pau, "A novel approach for dynamic traffic lights management based on Wireless Sensor Networks and multiple fuzzy logic controllers," *Expert Systems with Applications*, vol. 42, no. 13, pp. 5403–5415, 2015.

[40] M. B. Trabia, M. S. Kaseko, and M. Ande, "A two-stage fuzzy logic controller for traffic signals," *Transportation Research Part C: Emerging Technologies*, vol. 7, no. 6, pp. 353–367, 1999.

[41] Y. Zhang and Z. Ye, "Short-term traffic flow forecasting using fuzzy logic system methods," *Journal of Intelligent Transportation Systems*, vol. 12, no. 3, pp. 102–112, 2008.

[42] H. Wang, L. Zheng, and X. Meng, *Traffic Accidents Prediction Model Based on Fuzzy Logic*, Springer, Berlin, Germany, 2011.

[43] Y. Zhang and H. Ge, "Freeway travel time prediction using takagi-sugeno-kang fuzzy neural network," *Computer-aided Civil and Infrastructure Engineering*, vol. 28, no. 8, pp. 594–603, 2013.

[44] J. Zhao, "Research on prediction of traffic congestion state," in *Proceedings of the MATEC Web of Conferences*, Les Ulis, France, 2015.

[45] S. Sun, J. Chen, and J. Sun, "Traffic congestion prediction based on GPS trajectory data," *International Journal of Distributed Sensor Networks*, vol. 15, no. 5, Article ID 155014771984744, 2019.

[46] Y. Qi and S. Ishak, "A Hidden Markov Model for short term prediction of traffic conditions on freeways," *Transportation Research Part C: Emerging Technologies*, vol. 43, pp. 95–111, 2014.

[47] B. Jiang and Y. Fei, "Traffic and vehicle speed prediction with neural network and hidden markov model in vehicular networks," in *Proceedings of the 2015 IEEE Intelligent Vehicles Symposium (IV)*, pp. 1082–1087, IEEE, Seoul, 2015.

[48] G. Zhu, K. Song, P. Zhang, and L. Wang, "A traffic flow state transition model for urban road network based on Hidden Markov Model," *Neurocomputing*, vol. 214, pp. 567–574, 2016a.

[49] L. Zhu, R. Krishnan, F. Guo, J. Polak, and A. Sivakumar, "Early identification of recurrent congestion in heterogeneous urban traffic," in *Proceedings of the 2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, pp. 4392–4397, IEEE, Auckland, New Zealand, October 2019.

[50] Y. Xie, K. Zhao, Y. Sun, and D. Chen, "Gaussian processes for short-term traffic volume forecasting," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2165, no. 1, pp. 69–78, 2010.

[51] S. Jin, X. Qu, and D. Wang, "Assessment of expressway traffic safety using Gaussian mixture model based on time to collision," *International Journal of Computational Intelligence Systems*, vol. 4, no. 6, pp. 1122–1130, 2011.

[52] J. Jun, "Understanding the variability of speed distributions under mixed traffic conditions caused by holiday traffic," *Transportation Research Part C: Emerging Technologies*, vol. 18, no. 4, pp. 599–610, 2010.

[53] S. Shiliang, Z. Changshui, and Y. Guoqiang, "A bayesian network approach to traffic flow forecasting," *IEEE Transactions on Intelligent Transportation Systems*, vol. 7, no. 1, pp. 124–132, 2006.

[54] G. Asencio-Cortés, E. Florido, A. Troncoso, and F. Martínez-Álvarez, "A novel methodology to predict urban traffic congestion with ensemble learning," *Soft Computing*, vol. 20, no. 11, pp. 4205–4216, 2016.

[55] Z. Zhu, B. Peng, C. Xiong, and L. Zhang, "Short-term traffic flow prediction with linear conditional Gaussian Bayesian network," *Journal of Advanced Transportation*, vol. 50, no. 6, pp. 1111–1123, 2016.

[56] S. Wang, W. Huang, and H. K. Lo, "Traffic parameters estimation for signalized intersections based on combined shockwave analysis and Bayesian Network," *Transportation Research Part C: Emerging Technologies*, vol. 104, pp. 22–37, 2019.

[57] S. Wang, W. Huang, and H. K. Lo, "Combining shockwave analysis and Bayesian Network for traffic parameter estimation at signalized intersections considering queue spillback," *Transportation Research. Part C, Emerging Technologies*, vol. 120, Article ID 102807, 2020.

[58] X. Wang, K. An, L. Tang, and X. Chen, "Short term prediction of freeway exiting volume based on SVM and KNN," *International Journal of Transportation Science and Technology*, vol. 4, no. 3, pp. 337–352, 2015.

[59] Y. Liu, X. Feng, Q. Wang, H. Zhang, and X. Wang, "Prediction of urban road Congestion using a bayesian network approach," *Procedia—Social and Behavioral Sciences*, vol. 138, no. C, pp. 671–678, 2014.

[60] T. Ito and R. Kaneyasu, "Predicting traffic congestion using driver behavior," in *Proceedigs of the International Conference on Knowledge Based and Intelligent Information and Engineering Systems*, Marseille, France, 2017.

[61] K. Kumar, M. Parida, and V. K. Katiyar, "Short term traffic flow prediction for a non urban highway using artificial neural network," *Procedia - Social and Behavioral Sciences*, vol. 104, pp. 755–764, 2013.

[62] K. Kumar, M. Parida, and V. K. Katiyar, "Short term traffic flow prediction in heterogeneous condition using artificial neural network," *Transport*, vol. 30, no. 4, pp. 397–405, 2013.

[63] R. More, A. Mugal, S. Rajgure, R. B. Adhao, and V. K. Pachghare, "Road traffic prediction and congestion control using artificial neural networks," in *Proceedings of the 2016 International Conference on Computing, Analytics and Security Trends (CAST)*, pp. 52–57, IEEE, Pune, India, December 2016.

[64] C. Jacobé de Naurois, C. Bourdin, A. Stratulat, E. Diaz, and J.-L. Vercher, "Detection and prediction of driver drowsiness using artificial neural network models," *Accident Analysis & Prevention*, vol. 126, pp. 95–104, 2019.

[65] P. Kumar, S. P. Nigam, and N. Kumar, "Vehicular traffic noise modeling using artificial neural network approach," *Transportation Research Part C: Emerging Technologies*, vol. 40, pp. 111–122, 2014.

[66] V. Nourani, H. Gökçekuş, I. K. Umar, and H. Najafi, "An emotional artificial neural network for prediction of vehicular traffic noise," *Science of The Total Environment*, vol. 707, p. 136134, 2020.

[67] T. Alghamdi, K. Elgazzar, M. Bayoumi, T. Sharaf, and S. Shah, "Forecasting traffic congestion using arima modeling," in *Proceedings of the 2019 15th International Wireless Communications & Mobile Computing Conference (IWCMC)*, pp. 1227–1232, IEEE, Tangier, Morocco, October 2019.

[68] M. Chen, X. Yu, and Y. Liu, "PCNN: deep convolutional networks for short-term traffic congestion prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 11, pp. 3550–3559, 2018.

[69] C. Lee, Y. Kim, S. Jin et al., "A visual analytics system for exploring, monitoring, and forecasting road traffic Congestion," *IEEE Transactions on Visualization and Computer Graphics*, vol. 26, no. 11, pp. 3133–3146, 2020.

[70] H. Wang, L. Liu, S. Dong, Z. Qian, and H. Wei, "A novel work zone short-term vehicle-type specific traffic speed prediction model through the hybrid EMD-ARIMA framework," *Transportmetrica B: Transport Dynamics*, vol. 4, no. 3, pp. 159–186, 2016.

[71] X. F. Wang, X. Y. Zhang, Q. Y. Ding, and Z. Q. Sun, "Forecasting traffic volume with space-time ARIMA model," *Advanced Materials Research*, vol. 156-157, pp. 979–983, 2010.

[72] L. Kui-Lin, Z. Chun-Jie, and X. Jian-Min, "Short-term traffic flow prediction using a methodology based on ARIMA and RBF-ANN," in *Proceedings of the 2017 Chinese Automation Congress (CAC)*, pp. 2804–2807, Jinan, China, October 2017.

[73] Y. Liu and H. Wu, "Prediction of road traffic congestion based on random forest,"vol. 2, pp. 361–364, in *Proceedings of the 2017 10th International Symposium on Computational Intelligence and Design (ISCID)*, vol. 2, pp. 361–364, IEEE, Hangzhou, China, December 2017.

[74] W. Alajali, W. Zhou, S. Wen, and Y. Wang, "Intersection traffic prediction using decision tree models," *Symmetry*, vol. 10, no. 9, p. 386, 2018.

[75] 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 Generation Computer Systems*, vol. 104, pp. 142–158, 2020.

[76] X. Feng, X. Ling, H. Zheng, Z. Chen, and Y. Xu, "Adaptive multi-kernel SVM with spatial-temporal correlation for short-term traffic flow prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 6, pp. 2001–2013, 2019.

[77] S. Lu and Y. Liu, "Evaluation system for the sustainable development of urban transportation and ecological environment based on SVM," *Journal of Intelligent & Fuzzy Systems*, vol. 34, no. 2, pp. 831–838, 2018.

[78] Q. Shen, X. Ban, and C. Guo, "Urban traffic Congestion evaluation based on kernel the semi-supervised extreme learning machine," *Symmetry*, vol. 9, no. 5, p. 70, 2017.

[79] C. Yuan-Yuan, Y. Lv, Z. Li, and F.-Y. Wang, "Long short-term memory model for traffic congestion prediction with online open data," in *Proceedings of the 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, pp. 132–137, Rio de Janeiro, Brazil, December 2016.

[80] X. Ma, Z. Dai, Z. He, J. Ma, Y. Wang, and Y. Wang, "Learning traffic as images: a deep convolutional neural network for large-scale transportation network speed prediction," *Sensors*, vol. 17, no. 4, p. 818, 2017.

[81] R. Ke, W. Li, Z. Cui, and Y. Wang, "Two-stream multi-channel Convolutional neural network for multi-lane traffic speed prediction Considering traffic volume impact," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2674, no. 4, pp. 459–470, 2020.

[82] T. Bogaerts, A. D. Masegosa, J. S. Angarita-Zapata, E. Onieva, and P. Hellinckx, "A graph CNN-LSTM neural network for short and long-term traffic forecasting based on trajectory data," *Transportation Research Part C: Emerging Technologies*, vol. 112, pp. 62–77, 2020.

[83] T. Zhang, Y. Liu, Z. Cui, J. Leng, W. Xie, and L. Zhang, "Short-term traffic Congestion forecasting using attention-based long short-term memory recurrent neural network," in *Proceedings of the International Conference on Computational Science*, Faro, Portugal, June 2019.

[84] X. Di, Y. Xiao, C. Zhu, Y. Deng, Q. Zhao, and W. Rao, "Traffic congestion prediction by spatiotemporal propagation patterns," in *Proceedings of the 2019 20th IEEE International Conference on Mobile Data Management (MDM)*, pp. 298–303, Hong Kong, China, June 2019.

[85] X. Ma, H. Yu, Y. Wang, and Y. Wang, "Large-scale transportation network congestion evolution prediction using deep learning theory," *PloS One*, vol. 10, no. 3, Article ID e0119044, 2015.

[86] Z. Zhene, P. Hao, L. Lin et al., "Deep convolutional mesh RNN for urban traffic passenger flows prediction," in *Proceedings of the 2018 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*, pp. 1305–1310, IEEE, Guangzhou, China, October 2018.

[87] P. Li, M. Abdel-Aty, and J. Yuan, "Real-time crash risk prediction on arterials based on LSTM-CNN," *Accident Analysis & Prevention*, vol. 135, Article ID 105371, 2020.

[88] W. Xiangxue, X. Lunhui, C. Kaixun, X. Lunhui, C. Kaixun, and C. Kaixun, "Data-Driven short-term forecasting for urban road network traffic based on data processing and LSTM-RNN," *Arabian Journal for Science and Engineering*, vol. 44, no. 4, pp. 3043–3060, 2019.

[89] Q. Shen, X. Ban, C. Guo, and C. Wang, "Kernel based semi-supervised extreme learning machine and the application in traffic congestion evaluation," *Proceedings of ELM-2015*, vol. 1, pp. 227–236, 2016.

[90] Z. Zhang, A. Zhang, C. Sun et al., "Research on air traffic flow forecast based on ELM non-iterative algorithm," *Mobile Networks and Applications*, pp. 1–15, 2020.

[91] Y.-m. Xing, X.-j. Ban, and R. Liu, "A short-term traffic flow prediction method based on kernel extreme learning machine," in *Proceedings of the 2018 IEEE International Conference on Big Data and Smart Computing (BigComp)*, pp. 533–536, IEEE, Shanghai, China, January 2018.

[92] L. Lin, J. C. Handley, and A. W. Sadek, "Interval prediction of short-term traffic volume based on extreme learning machine and particle swarm optimization," in *Proceedings of the 96th Transportation Research Board Annual*, Washington DC, USA, 2017.

[93] S. Zhang, Y. Yao, J. Hu, Y. Zhao, S. Li, and J. Hu, "Deep autoencoder neural networks for short-term traffic Congestion prediction of transportation networks," *Sensors*, vol. 19, no. 10, p. 2229, 2019.

[94] D. Huang, Z. Deng, S. Wan, B. Mi, and Y. Liu, "Identification and prediction of urban traffic congestion via cyber-physical link optimization," *IEEE Access*, vol. 6, pp. 63268–63278, 2018.

[95] V. Ahsani, M. Amin-Naseri, S. Knickerbocker, and A. Sharma, "Quantitative analysis of probe data characteristics: coverage, speed bias and congestion detection precision," *Journal of Intelligent Transportation Systems*, vol. 23, no. 2, pp. 103–119, 2019.

[96] S. J. Kwon, *Artificial Neural Networks*, Nova Science Publishers, New York, NY, USA, 2011.

[97] Y. Liu, Z. Liu, and R. Jia, "DeepPF: a deep learning based architecture for metro passenger flow prediction," *Transportation Research Part C: Emerging Technologies*, vol. 101, pp. 18–34, 2019.