**TRAFFIC MANAGEMENT SYSTEM**

**Integrating Historical Traffic Data And Machine Learning Algorithms To Predict Congestion Patterns**.

**Abstract:**

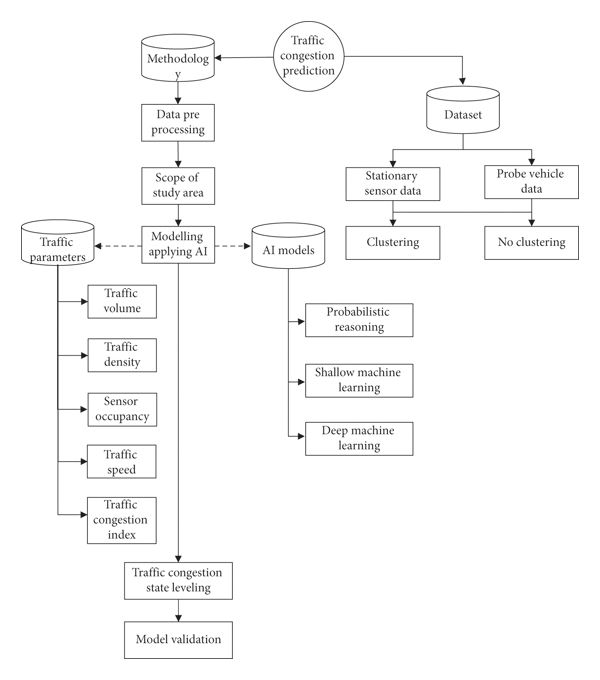
In recent years, traffic congestion prediction has led to a growing research area, especially of machine learning of artificial intelligence (AI). With the introduction of big data by stationary sensors or probe vehicle data and the development of new AI models in the last few decades, this research 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 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.
* A general layout of the prediction approaches is provided in Section [2](https://www.hindawi.com/journals/jat/2021/8878011/#sec2). The data collection sources and congestion forecasting models are explained in Sections [3](https://www.hindawi.com/journals/jat/2021/8878011/#sec3)–[6](https://www.hindawi.com/journals/jat/2021/8878011/#sec6) 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..

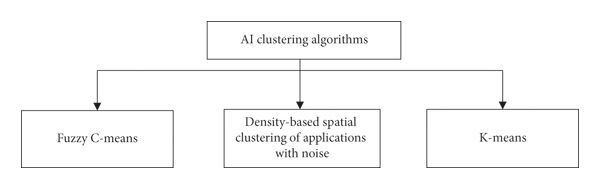
[](https://www.hindawi.com/journals/jat/2021/8878011/fig1/)

#### 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.

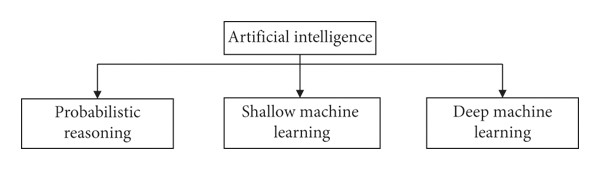
##### **3.1. Clustering Algorithms:**

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](https://www.hindawi.com/journals/jat/2021/8878011/fig2/) shows the commonly used AI clustering models in this field of research. The models are described briefly in this section.

[[](https://www.hindawi.com/journals/jat/2021/8878011/fig2/)](https://www.hindawi.com/journals/jat/2021/8878011/fig2/" \t "_blank)

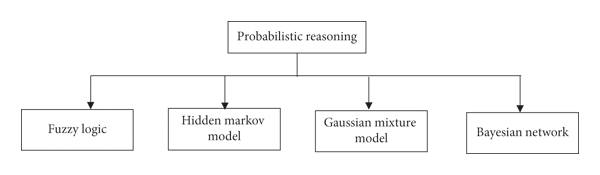
**4. Applied Methodology:**

* 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](https://www.hindawi.com/journals/jat/2021/8878011/fig3/) shows the main branches—probabilistic reasoning and machine learning (ML).

[[](https://www.hindawi.com/journals/jat/2021/8878011/fig3/)](https://www.hindawi.com/journals/jat/2021/8878011/fig3/" \t "_blank)

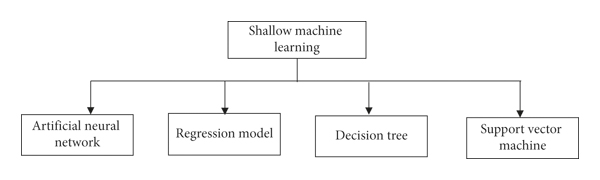
##### **4.1. Probabilistic Reasoning**

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](https://www.hindawi.com/journals/jat/2021/8878011/fig4/).

[[](https://www.hindawi.com/journals/jat/2021/8878011/fig4/)](https://www.hindawi.com/journals/jat/2021/8878011/fig4/" \t "_blank)

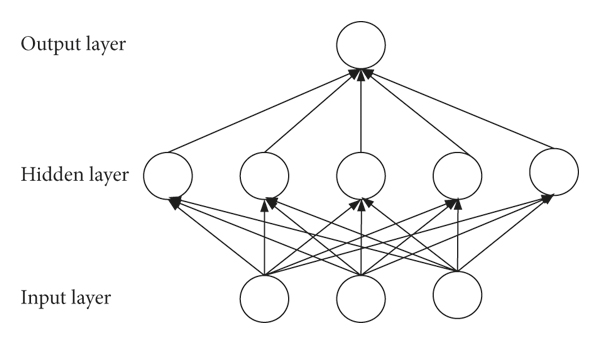
##### **4.2. Shallow Machine Learning:**

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](https://www.hindawi.com/journals/jat/2021/8878011/fig6/).

[[](https://www.hindawi.com/journals/jat/2021/8878011/fig6/)](https://www.hindawi.com/journals/jat/2021/8878011/fig6/" \t "_blank)

###### **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](https://www.hindawi.com/journals/jat/2021/8878011/fig7/) 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.

[[](https://www.hindawi.com/journals/jat/2021/8878011/fig7/)](https://www.hindawi.com/journals/jat/2021/8878011/fig7/" \t "_blank)

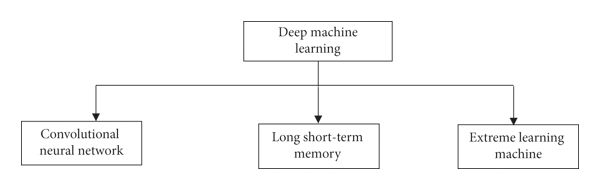
* Unlike the previous studies, those focused on traffic flow parameters to conduct traffic congestion prediction research; Ito and Kaneyasu [60] analysed drivers’ behaviour in predicting congestion. They showed that vehicle operators act differently on different phases of the journey. They used one layered 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.

###### **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.
* 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 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].

##### **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](https://www.hindawi.com/journals/jat/2021/8878011/fig8/) and discussed in this section.

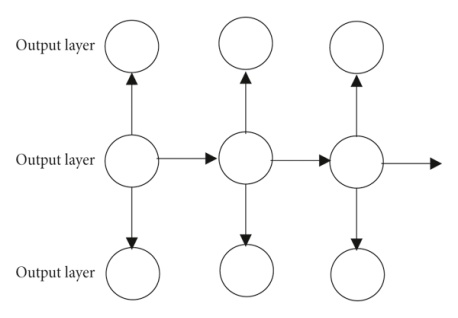
[[](https://www.hindawi.com/journals/jat/2021/8878011/fig8/)](https://www.hindawi.com/journals/jat/2021/8878011/fig8/" \t "_blank)

###### **4.3.1. Convolution Neural Network:**

Convolution 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.

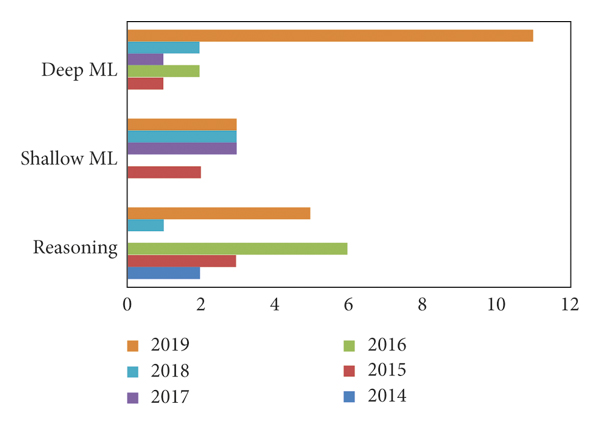
**4.3.2. Recurrent Neural Network**

* Recurrent neural network (RNN) has a wide usage in the sequential traffic data processing by considering the influence of the related neighbour (Figure [9](https://www.hindawi.com/journals/jat/2021/8878011/fig9/)). 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.

[[](https://www.hindawi.com/journals/jat/2021/8878011/fig9/)](https://www.hindawi.com/journals/jat/2021/8878011/fig9/" \t "_blank)

#### 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.
* 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](https://www.hindawi.com/journals/jat/2021/8878011/fig10/)).

[[](https://www.hindawi.com/journals/jat/2021/8878011/fig10/)](https://www.hindawi.com/journals/jat/2021/8878011/fig10/" \t "_blank)

* 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](https://www.hindawi.com/journals/jat/2021/8878011/tab4/), [5](https://www.hindawi.com/journals/jat/2021/8878011/tab5/)–[6](https://www.hindawi.com/journals/jat/2021/8878011/tab6/) summarize the advantage and weaknesses of the algorithms of different branches.

**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.

#### 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. 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.

**Traffic congestion is a global problem that many cities suffering from it.**