Tikrit Journal of Pure Science Vol. 28 (3) 2023

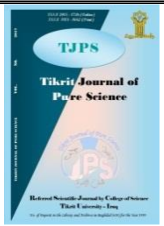
TJPS

https://doi.org/10.25130/tjps.v28i3.1428

Tikrit Journal of Pure Science

ISSN: 1813 – 1662 (Print) --- E-ISSN: 2415 – 1726 (Online)

Journal Homepage: http://tjps.tu.edu.iq/index.php/j



Using a new algorithm in Machine learning Approaches to estimate level-of-service in hourly traffic flow data in vehicular ad hoc networks

Ahmed Ibrahim Turki[1], Saad Talib Hasson[2]

[1]*Department of Physics, College of Education, University of Samarra, Samarra, Iraq*

[2]*Information Networks Department, University of Babylon, Babylon, Iraq*

ARTICLE INFO.

Article history:

-Received:	**13 / 12 / 2022**
-Received in revised form:	**22 / 2 / 2023**
-Accepted:	**14 / 3 / 2023**
-Final Proofreading:	**15 / 5 / 2023**
-Available online:	**25 / 6 / 2023**

Keywords: Traffic flow data; Level-of-Service; Random Forest; SVM, KNN, VANET

Corresponding Author:

Name: Ahmed Ibrahim Turki

E-mail:

ahmed.ibrahim@uosamarra.edu.iq

saad_aljebori@itnet.uobabylon.edu.iq

Tel:

ABSTRACT

The primary goals of transportation agencies and researchers studying traffic operations are to ease traffic and increase road safety through the use of vehicular ad hoc networks. Agencies can't achieve their goals without reliable and consistent data on the current traffic situation. The Level-of-Service (LOS) index is a helpful measure of freeway traffic operations. Conventional fixed-location cameras and sensors are impractical and expensive for gathering reliable traffic density data on every road in large networks. Flow data is a new, low-cost option that has the potential to boost safety and operations. This study proposes an algorithm for hourly LOS assessment by incorporating flow data provided by the MIDAS (Motorway Incident Detection and Automatic Signaling) system. The proposed algorithm uses machine learning techniques to classify LOS data based on the flow of traffic. The input features that are subject to prediction are a group of technical indicators. The real-world LOS was determined by analyzing data from stationary sensors. The outcomes demonstrate that technical indicators can be utilized to enhance the accuracy of LOS estimation (Random Forest= 93.1, k-nearest neighbors = 92.5, and Support Vector Machine = 91.4). The current work introduces a novel approach to the selection of technical indicators and their use as features, which allows for highly accurate short-term prediction of LOS estimation.

استخدام خوارزمية جديدة في مناهج التعلم الآلي لتقدير مستوى الخدمة في بيانات تدفق حركة المرور بالساعة في الشبكات المخصصة للمركبات

احمد إبراهيم تركي [1], سعد طالب حسون [2]

[1] *قسم الفيزياء ، كلية التربية ، جامعة سامراء ، سامراء، العراق*

[2] *قسم شبكات المعلومات ، جامعة بابل ، بابل ، العراق*

الملخص

تتمثل الأهداف الأساسية لوكالات النقل والباحثين الذين يدرسون عمليات المرور في تسهيل حركة المرور وزيادة السلامة على الطرق من خلال استخدام شبكات مخصصة للمركبات (VANET). لا تستطيع الوكالات تحقيق أهدافها بدون بيانات موثوقة ومتسقة حول الوضع المروري الحالي. مؤشر مستوى الخدمة (LOS) هو مقياس مفيد لعمليات حركة المرور على الطرق السريعة. تعد أجهزة الاستشعار والكاميرات التقليدية الخاصة بالمواقع الثابتة باهظة الثمن وغير عملية لجمع بيانات الكثافة الموثوقة على كل طريق في الشبكات الكبيرة. تعد بيانات التدفق خيارًا جديدًا منخفض

https://doi.org/10.25130/tjps.v28i3.1428

التكلفة لديه القدرة على تعزيز السلامة والعمليات. تقترح هذه الدراسة خوارزمية لتقييم LOS كل ساعة من خلال دمج بيانات التدفق التي يوفرها نظام MIDAS (اكتشاف حوادث الطرق السريعة والإشارات التلقائية). تستخدم الخوارزمية المقترحة نماذج التعلم الآلي لتصنيف بيانات LOS بناءً على تدفق حركة المرور. ميزات الإدخال التي تخضع للتنبؤ هي مجموعة من المؤشرات الفنية. تم تحديد LOS في العالم الحقيقي من خلال تحليل البيانات من أجهزة الاستشعار الثابتة. توضح النتائج أنه يمكن استخدام المؤشرات الفنية لتعزيز دقة تقدير LOS (الغابة العشوائية = 93.1, آلة متجه الدعم = 91.4, أقرب جيران–k= 92.5). يقدم العمل الحالي نهجًا جديدًا لاختيار المؤشرات الفنية واستخدامها كميزات، مما يسمح بالتنبؤ الدقيق للغاية على المدى القصير لتقدير مستوى الخدمة.

1- Introduction

In a vehicular ad hoc network (VANET), traffic conditions can't be accurately assessed without the help of intelligent transportation systems (ITS). Road work planning, traffic operations, congestion management, and assessing traffic queues are just a few of the uses for ITS traffic measurements. For the purpose of estimating traffic performance and state, the Highway Capacity Manual (HCM) defines six LOS. The HCM offers formulas for calculating LOS based on traffic volume and road conditions [1]. An important part of LOS evaluation is the speed, flow, and density of the traffic [1, 2, 3]. The transportation agencies often require hourly data on the traffic situation and LOS for various stretches of freeway, either in real time or historically fixed location sensors like remote traffic microwave sensors (RTMS), loop detectors, laser sensors, magnetic sensors, license plate recognition (LPR), and video image systems [4,5] have long been used to collect traffic data (travel time, speed, density, and flow). Data collection techniques that rely on stationary nodes are notoriously costly and space-consuming. Recently, data-driven ITS has resulted in multi-source, high-performance, and potent solutions for transportation systems [6]. The use of "probe vehicles" and "floating cars" for data collection has recently received a lot of attention. These strategies collect information about traffic through the use of cutting-edge technologies like connected vehicles (CVs), Wi-Fi, Bluetooth sensors, cellular networks and smartphones [7,8]. These tools not only open up new possibilities for collecting crowdsourced data, but also produce valuable information that can be used in a variety of transportation analyses, including those concerned with traffic safety [9,10,11,12], public transit [13,14], and energy consumption and emissions [15,16]. Big data is being used in the transportation sector to propose novel ideas and solutions that have not been explored before [17]. Predicted traffic flows are a key input into LOS calculations for highways, and as a result, they can help drivers and passengers make more informed decisions about which routes to take. Knowing "when and where" congestion will occur is helpful for transportation planning because it allows experts to allocate resources to the roads at risk of congestion, which can reduce traffic congestion over time. To that end, traffic flow prediction [18] [19] [20] has become a hot topic in recent years as a means to estimate LOS due to its substantial advantages over other devices.

Since VANET uses traffic flow data, many city governments and departments of transportation (DOTs) have made deals with data providers like MIDAS to work together. Flow data has been used in many different ways, such as to measure performance and find problems. The focus of this paper is on MIDAS. In the UK, the MIDAS system is made up of a network of traffic sensors, mostly inductive loops, that send information about traffic volumes and average speeds to a regional control center (RCC). The RCC can then change variable message signs and advisory speed limits automatically. When flow data is collected, it gives us a chance to come up with a new way to measure LOS based on the features and characteristics of the data. This study comes up with a new way to measure LOS on freeways in VANET that uses technical indicators from flow data. With this method, you don't need fixed traffic volume sensors to make new tools for LOS assessment and hourly traffic status data on freeways. The proposed method could be thought of as an addition to the traditional HCM LOS calculation method, which is based on the amount and speed of traffic. Here’s how the rest of this paper is put together: In the next section, "Methodology," the traditional way to figure out LOS and the proposed way to figure it out are shown. In this section, we also talk about some methods for data mining. Then, the data used in this study are talked about, and then the results of using the methodology are given. At the end of the paper, suggestions are made for further research.

1- Related Works

This part reviews the most relevant literature pertaining to this study, summarizes traffic status and LOS assessment methods, and discusses the research gaps. Studies have typically relied on single or multiple parameters, such as traffic flow [21], traffic speed [22], and traffic density [23], to explain traffic status and LOS. Previous research has relied on a wide variety of methods and data sets, including sensor readings [24], probe vehicles [25], camera videos and images [26], CVs [2], and simulation environments [2][23]. Regarding approach, statistical modeling [23], artificial neural networks [24,25], Kalman filters (KF) [25], image processing [27], and machine learning (ML) [21,26] have all seen extensive use. Table 1 presents the most relevant studies that have proposed alternative methods for LOS assessment.

Table 1: is a summary of the different ways that level-of-service (LOS) can be measured.

No.	*Reference*	*Year*	*Data*	*Index used*	*Method*
1	[2]	2017	Simulation (speed, density)	- Average speed - CV penetration rate	- Artificial intelligence
2	[22]	2016	Floating Car Data (speed)	- Average speed	- Speed threshold
3	[28]	2008	Sensor data (speed, density, travel time)	- Travel speed range - Most restrictive condition - Value of travel time	- Travel time reliability threshold
4	[29]	2019	Wi-Fi probe vehicle (Speed, travel time)	- Planning Time Index - Buffer Time Index - Travel Time Index	- Travel time reliability threshold - Statistical regression
5	[30]	2019	Travel time data provided by North Carolina DOT	- Planning Time Index - Buffer Time Index - Average travel time	- Travel time reliability threshold - Regression model
6	[31]	2020	Simulation (travel time)	- Planning Time Index - Buffer Time Index	- Travel time reliability threshold - Statistical regression

As we've already talked about, the most attention in the past literature was paid to HCM density-based LOS. Some studies also used travel time and speed changes to figure out LOS. No data on traffic flow has been used to figure out LOS. This study fills a gap in integrating flow data for LOS assessment with the help of technical indicators as features. The results of this study can help agencies figure out LOS for different segments without having to install new fixed location equipment.

- To meet the goals of the study and estimate hourly LOS based on flow data, the following machine learning classification methods were used:

1- Random Forest (RF): RF is a classification technique that uses a collection of random decision trees to make a more accurate prediction than using either one alone. Here, each tree is constructed separately from the others. The data is then classified using a majority vote across all trees, with Gini impurity serving as the function to measure the quality of the split at each node [32]. The Gini impureness at a given node N is defined as:

$$\boldsymbol{G(N) = 1 - (P_1)^2 - (P_{-1})^2} \qquad (10)$$

where *Pi* is the proportion of the population with class label i.

2- Support Vector Machines (SVM): Support vector machines are a famous classification technique that uses margins. The SVM algorithm determines, for each class, the ideal SVM that provides the greatest distance to other classes. The algorithm delineates boundaries and assigns classes to data by computing optimal support vectors [33].

3- K-Nearest Neighbors (KNN): The use of KNNs, a non-parametric technique, in the classification process is commonplace. In this approach, the entire set of training data is mapped onto a feature space with n dimensions (where n is the number of input features). The algorithm takes the Euclidean distance between each observation and its nearest neighbors and finds the k closest neighbors. After that, it determines a label based on how often it appears among the neighbors [34].

3-Materials and Methods

In this paper, a method based on traffic flow data to determine the hourly level of service-based traffic status are used. This approach takes into account the volume of traffic on a given stretch of road in order to determine the technical indicators that characterize the state of traffic along that route. The study, which will be detailed below, relies heavily on data from the MIDAS traffic flow. This section elaborates on the proposed algorithm from this research. The various stages of the proposed method are as follows, as depicted in the research framework (Figure 1):

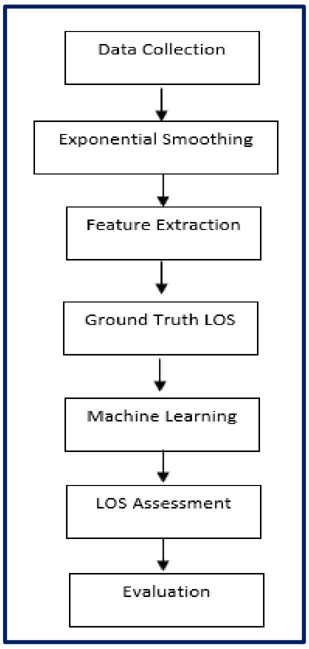


Fig. 1: framework of supervised learning in the current work

https://doi.org/10.25130/tjps.v28i3.1428

3.1 Data collection

Massive amounts of data are continuously computed by MIDAS. Storing MIDAS travel time and traffic data is the first step in conducting such an investigation. Data on traffic volumes and travel times were recorded at 15-minute intervals thanks to a Python code. Using raw data from the real world always comes with the risk of encountering problems like noise and missing values. The data was cleaned and checked for errors before being used. As far as possible, missing values and outliers were removed or identified. The next step was to gather MIDAS traffic data in order to determine the hourly flow and ground truth for the level of service. To evaluate the efficacy of the algorithm [35], MIDAS data of the M25 highway between Junction (13-14) in the United Kingdom's busiest highway was collected, as shown in Figure 2.

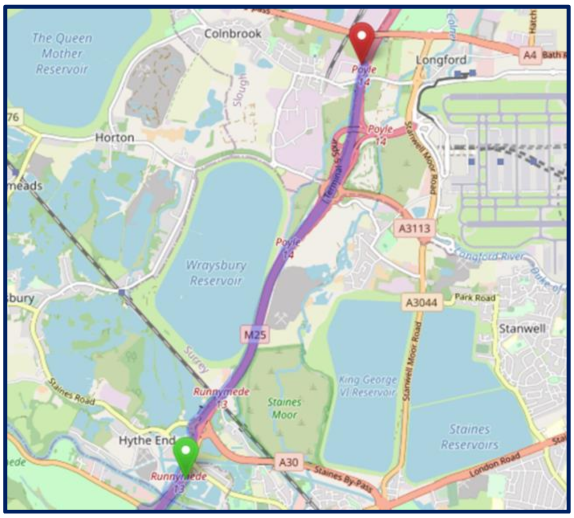


Fig. 2: Part of the M25 highway chosen for the study from Open Street Map

3.2 Exponential Smoothing

With exponential smoothing, more importance is placed on more recent observations, while less important observations from further back in time are given weights that decrease at an exponential rate. Recursively finding the exponentially smoothed statistic of a series *Y* looks like this:

$S_0 = Y_0$

$for\ t > 0;\ S_t = \alpha * Y_t + (1-\alpha) * S_{t-1}$ (1)

where α represents a smoothing factor. Increasing the friction has the opposite effect, increasing the roughness. α = 1, so the smoothed statistic is identical to the raw data. When multiple consecutive observations are available, the smoothed statistic St can be computed. Through this process of smoothing, the model is better able to detect the long-term trend in the behavior of traffic flows by eliminating the effect of random variation or noise in the underlying data. Following the exponential smoothing of the time series data, a feature matrix is constructed from which technical indicators are derived.

3.3 Feature extraction from data

The only variables considered are vehicle travel time and traffic flow over the course of several days. As a result, the format can be used to evaluate our input data (date, traffic flow). These indicators are derived from the data:

- **Average True Range (ATR):** The ATR measures the deviation from the average over a given time period [36] and the size of the range over that time period. It is formulated as per Eq. (2). The true range is indicated here by the symbol TR.

$ATR_n = \frac{1}{n}\sum_{i=1}^{n} TR_i$ (2)

Where:

$TR_i = MAX\{A_n, B_n, C_n\}$ (3)

$A_n = HighestFlow_n - LowestFlow_n$

$B_n = |HighestFlow_n - Flow_n|$

$C_n = |LowestFlow_n - Flow_n|$

- **Simple Moving Average (SMA):** Adding the traffic volume of a vehicle fleet over a range of times and then dividing by the range of times yields the SMA [37].

$SMA_n = \frac{1}{n}\sum_{i=1}^{n-1} Flow_{t-i}$ (4)

- **Exponential Moving Average (EMA):** EMA is an abbreviation for "exponential moving average" [38]. Using Eq. (5), where R is the traffic flow for the most recent period, D is the smoothing constant equal to 2/(nu+1), nu is the number of traffic flows in the SMA estimated by EMA, and EMAu is the EMA for traffic flows in the past, by using the equation 5 below:

$EMA = (R - EMA_u) * D + EMA_u$ (5)

- **Relative strength index (RSI):** The normalized current flow is a percentage between (0-100). The name of this oscillator is deceptive because it does not make comparisons between instruments; rather, it depicts the current flow in terms of how it compares to pieces that have been produced within the chosen lookback window length [32]. The equation for the RSI is (6).

$RSI_n = 100 - \left[\frac{100}{D_n}\right]$ (6)

Where:

$D_n = \left[1 - \frac{\frac{1}{n}\sum_{i=1}^{n} Flowup[flow_i - flow_n]}{\frac{1}{n}\sum_{i=1}^{n} flowdown[flow_i - flow_n]}\right]$ (7)

- **Rate of Change (ROC):** The ratio of the current flow to the average flow over the window used to measure the time period under observation [33] is a technical indicator that measures the relative magnitude of the two flows. This is the formula for determining the ROC:

$\mathrm{ROC} = (\text{Current flow / flow of n bars ago}) - 1.0) * 100$ (8)

- **Momentum (MOM):** Using data from a predetermined number of periods in the past, the MOM indicator evaluates how the current flow compares to that data. Akin to the "Rate of Change" indicator, the MOM does not normalize the flow, resulting in different indicator values for various

https://doi.org/10.25130/tjps.v28i3.1428

instruments depending on their point values [39]. Equation (9) is used to determine the MOM:

$$\text{MOM} = \text{Current flow} - \text{Flow of n periods ago} \quad (9)$$

3.4 Ground Truth LOS:

Level of service (LOS) is a popular metric for gauging how well a given stretch of road is performing. With data from flow and road characteristics, the HCM classified freeways and highways into six LOS groups. For highway sections, HCM uses traffic volume as the primary LOS metric [2]. Each LOS's flow is detailed in Table 2 [1]. In this investigation, the LOS was determined hourly based on traffic volume collected by MIDAS sensors. The hourly LOS was calculated using the traffic volume from Table 2. The LOS that was computed was used as the standard of comparison. The LOS model presented below makes use of hourly input data that was labeled with ground truth values.

Table 2: Description of various LOS derived from the HCM [1].

LOS	Flow (veh/hour/lane)	Description
A	Under 700	Free flow
B	700-1,100	Reasonably free flow
C	1,100-1,550	Stable flow (acceptable delays)
D	1,550-1,850	When flows are increased, speeds decreased marginally.
E	1,850-2,200	The state of being close to or at full capacity in operation
F	Over 2200	Breakdown flow

3.5 Machine Learning Methods

Several machine learning algorithms were put to the test in this study. The three most accurate methods (Random Forest, Support Vector Machines, K-nearest Neighbor, Decision Tree, Boosted Tree, Nave Bayes, and Multinomial Logistic Regression) were chosen from a group of seven (Random Forest, Support Vector Machines, K-nearest Neighbor, Decision Tree, Boosted Tree, Nave Bayes, and Multinomial Logistic Regression). Different machine learning methods were used in this study, so they had to be compared to find the best one. The preferred model and features were chosen based on classification accuracy, recall, precision, f-score, and support. In this study, the ratio of correctly labeled predictions (LOS) to ground truth data is measured by accuracy.

4- Experiments and Results

As a first step, this section supplies summary statistics for all of the data sources used. Next, the findings of the ML models are shown. All of the analyses and visuals in this section were created using the Python programming language. The datasets also did not contain any missing values that represented more than one percent of the entire population. When determining traffic flow data for the M25, factors such as profile diversity, profile reputation, and profile geometry validity were considered. The efficiency of the model allows for a range of values, which were taken into account while simulating traffic on the busiest highway. Additionally, the model's robustness should be assessed. Naturally, it would be easier to predict the free flow or breakdown of traffic that is relatively stable than traffic that is relatively noisy. From an engineering perspective, less variation in the data accounts for stability and means that ML classifiers can make more accurate predictions. Accuracy, recall (also known as sensitivity), precision, and f-score are the performance metrics used to assess the stability of a multiclass classifier.

4.1. Using Machine Learning for LOS Classification

This research used three different machine learning models to categorize LOS. This paper reports that KNN, SVM, and RF achieved the highest accuracy rates of all the methods tried. To get rid of unexpected local variation, exponential smoothing was used in this work. Figure 3 shows that, compared to the previously used classifiers, the results from RF, KNN, and SVM are superior. To achieve this goal, each technique used a grid of hyperparameter values with varying values to tune hyperparameters and choose the best model, as described in this Section. Table 3 displays the results of LOS classification using data from the M25 highway. Table 3 lists the various performance metrics used to assess the reliability of a multiclass classifier.

Table 3: Summary of classification ML methods

Classifier	LOS	Accuracy	Precision	Recall	F1-score	Support
RF	A	93.16	1.00	1.00	1.00	44
	B		1.00	1.00	1.00	9
	C		0.85	0.92	0.88	25
	D		0.88	0.88	0.88	34
	E		1.00	0.60	0.75	5
SVM	A	91.4	1.00	1.00	1.00	44
	B		1.00	1.00	1.00	9
	C		0.81	0.88	0.84	25
	D		0.85	0.85	0.85	34
	E		1.00	0.60	0.75	5
KNN	A	92.55	1.00	1.00	1.00	34
	B		1.00	1.00	1.00	7
	C		0.78	0.95	0.86	19
	D		0.93	0.83	0.88	30
	E		1.00	0.75	0.86	4

This study's results show that machine learning techniques can be used to figure out LOS. The machine learning techniques used got results of 93.16, 91.4, and 92.55, which is good for a classification with six categories. The RF did the best out of the machine learning techniques that were chosen. For the best model (selected RF), the hyperparameters were 300 trees, a maximum of 2 features, and a maximum tree depth of 3.

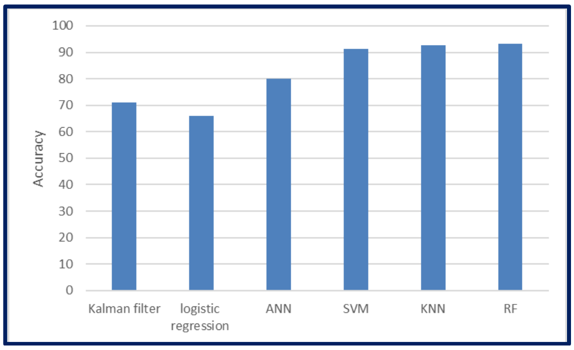


Fig. 3: Comparison of the accuracy achieved with different classifiers.

4.2. Sensitivity Analysis

Finally, the importance of each technical indicator was investigated via sensitivity analysis applied to the hourly random forest model. To identify the input parameters that most affect robustness and model performance, a sensitivity analysis is conducted [40]. To conduct this study, each technical indicator was first removed once from the model input before checking its accuracy. Since the factors are swapped out and the model is reevaluated after each iteration, this method is known as a parametric bootstrap [41]. The results of each eliminated technical indicator are summarized in Table 4. Not extracting any indicator from the sensitivity test gave a very high accuracy (93.16).

By the looks of things, the SMA was the most important technical indicator, with a drastic drop in model accuracy (accuracy = 87.28) after its removal. Once ATR was taken out of the model, the accuracy was very close to the original (accuracy = 93.11), making it the least important parameter. Even though the accuracy has increased with MOM, the overall profile of the results is not quite as good as when using all the technical indicators, so MOM also has low significance. Based on these results, it seems that the SMA is a reliable technical indicator for LOS estimation.

Table 4: Random Forest hourly sensitivity analysis

Parameter removed	Accuracy
None	93.16
ATR	93.15
SMA	88.88
EMA	87.28
RSI	89.74
ROC	92.3
MOM	93.1

4.3. Feature importance

It is hypothesized in this research that LOS classification accuracy can be enhanced by using data from technical indicators. Thus, the chosen RF model was used as the basis for an importance analysis of the variables involved (Fig. 4). As a result of using the RF model, it was possible to calculate an average statistically significant decrease in the Gini index. The significance of a variable is better captured by a higher value of this index. Significantly more weight is given to the SMA and EMA of traffic flow when calculating LOS. Figure 4 shows that ROC, MOM, and ATR are all less significant in the classification of LOS.

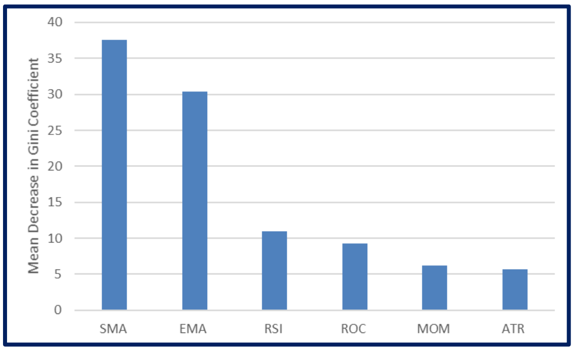


Fig. 4: Feature importance plot.

In order to better predict LOS, this research proposes a new method that incorporates traffic flow data and machine learning algorithms. However, this research was not without its flaws. The methodology's inherent sensitivity to factors like speed and weather conditions was ignored in this investigation. However, spatial flow variation was disregarded. Potentially useful in assessing LOS in the future is spatial variation, which can be gathered through further study. It is possible to account for the difference in flow between the upstream and downstream sections when estimating LOS. Deep learning and other sophisticated approaches can be used for this purpose. In the future, researchers may be able to capture spatial and temporal variation in the same study by employing deep neural networks like convolutional neural networks (CNNs) and recurrent neural networks (RNNs). By analyzing whether or not including TI characteristics improved traffic flow forecasting accuracy, this article assessed TIs' explanatory power. In general, it has been shown that TIs may capture the effects of behavioral biases in traffic flow, resulting in significantly lower prediction errors when using ML models. Our research showed that ATR, SMA, EMA, RSI, ROC, and MOM are the most effective TIs for predicting traffic volumes. In particular, our findings recommend including these TIs into the proposed ML models. Evidence has been found that TI performance varies by model; However, both SMA and EMA improved the accuracy of the ML models from 88.88% and 87.28% to 93.16%, respectively.

5- Conclusion

Rapidly expanding quantities of data on traffic flows in VANET are now available, and machine learning provides a means of analyzing them. This research provided a fresh approach to using flow information in LOS evaluation. The UK's M25 freeway between junctions 13 and 14 was used for the experiment. Six input metrics (ATR, SMA, EMA, RSI, and ROC MOM) were generated based on the acquired MIDAS traffic data. LOS was classified hourly using machine learning algorithms. The traffic density and LOS ground truth were estimated using HCM density criteria, with both calculated using data received from fixed-position loop sensors. Using a combination of machine learning and data on traffic flows, this study shows that level of service in VANET can be estimated with some degree of accuracy. The outcomes demonstrated that incorporating technical indicators as input can considerably raise the accuracy of the model. And when compared to other classification approaches on training datasets, RF performed the best of all (accuracy = 93.16 percent). It was concluded that this study will encourage others to investigate the possibilities of technical feature engineering because this is the first study to apply technical indicators features to level of service predictions in vehicular networks.

Although it has been considered that some of the most common and basic technical analysis indicators can explain phenomena, more advanced technical analysis indicators may be better at making predictions, and this is an area that future study may focus on.

Appendix A (Hyperparameters)

The hyperparameters of the optimal model are listed in Table 4 below. It should be mentioned that scikit-learn [42] is the tool we use to implement machine learning models.

Table 4: Grid-search selected model hyperparameters.

Classifier Models	Hyperparameters
RF	n_estimators=300, criterion= gini, random_state = 0
KNN	n_neighbors=5
SVM	C=1.0, kernel='linear'

Appendix B: MIDAS Dataset

Since April 2015, Highway England (HE), which is in charge of all the motorways and category "A" roads in England, has sent information about them every 15 minutes. This is called the "Strategic Road Network" in England. Major roads in category "A" are freeways, roads with two lanes, and motorways. Every minute, a record is made of the Motorway Incident Detection and Automatic Signaling (MIDAS) original gold dataset. It had rules about how the data collected at the site should be recorded. The most important ones are: publication time, speed (threshold: 240 km/h), vehicle flows (threshold: 120 veh/min), occupancy, and headway are reported per lane. Traffic monitoring equipment on the side of the road divides the flow of vehicles into five groups based on the length of each vehicle. These sorted flows of vehicles were converted to the volumetric unit of vehicles per minute for each lane and then added together to get readings for the carriageway [43]. The important data fields in the MIDAS traffic flow dataset are shown in Table 5 below. Each model site's files are made every month. Since Highway England is in charge of all the major highways, junctions, and motorways, each file only has flow, speed, and day logs from those places.

Table 5: shows the flow of traffic, along with other field names and descriptions that are only found in the MIDAS Dataset.

MIDAS ID	An identifier unique to the NTIS link.
Legacy MIDAS ID	An identifier unique to the NTIS link.
Site Name	A description of the site.
Local Date	Date local to BST.
Local Time	15-minute time intervals local to BST.
Day Type	The following are valid: • 0 - First working day of normal week; • 1 - Normal working Tuesday; • 2 - Normal working Wednesday; • 3 - Normal working Thursday; • 4 - Last working day of normal week; • 5 - Saturday, but excluding days falling within type 14; • 6 - Sunday, but excluding days falling within type 14; • 7 - First day of school holidays; • 9 - Middle of week - school holidays, but excluding days falling within type 12, 13 or 14; • 11 - Last day of week - school holidays, but excluding days falling within type 12,13 or 14; • 12 - Bank Holidays, including Good Friday, but excluding days falling within type 14; • 13 - Christmas period holidays between Christmas day and New Year’s Day; • 14 - Christmas Day/New Year’s Day.
Total Carriageway Flow	The number of vehicles detected on any lane within the 15-minute time slice.
Total Flow vehicles less than 5.2m	The number of vehicles less than 5.2m detected on any lane within the 15-minute time slice.
Total Flow vehicles 5.21m - 6.6m	Number of vehicles between 5.21m - 6.6m detected on any lane within the 15-minute time slice.
Total Flow vehicles 6.61m - 11.6m	The number of vehicles between 6.61m - 11.6mn detected on any lane within the 15-minute time slice.
Total Flow vehicles above 11.6m	The Number of vehicles above 11.6m detected on any lane within the 15-minute time slice.
Speed Value	The average speed in km/h. of all vehicles for all lanes measured by the site over the 15-minute period.
Quality Index	The Indication of the quality of the data provided. The number of valid one-minute records reported and used to generate the Total Traffic Flow and speed. A quality index of 0 indicates no valid records.

References

[1] Mahdi, M. A., & Hasson, S. T. (2018). Complex agent network approach to model mobility and connectivity in vehicular social networks. *J. Eng. Appl. Sci*, *13*, 2288-2295.

[2] Khan, S. M., Dey, K. C., & Chowdhury, M. (2017). Real-time traffic state estimation with connected vehicles. *IEEE Transactions on Intelligent Transportation Systems*, *18*(7), 1687-1699.

[3] Hernandez, S., Tok, A., & Ritchie, S. G. (2013). Density Estimation using Inductive Loop Signature based Vehicle Re-identification and Classification.

[4] Abdullah, M. Y., & Kanoosh, H. M. (2012). Voronoi Based Coverage Control for Wireless Sensors Networks. *Tikrit Journal of Pure Science*, *17*(3).

[5] Mahdi, M. A., & Hasson, S. T. (2017). Grouping vehicles in vehicular social networks. *Kurdistan Journal of Applied Research*, *2*(3), 218-225.

[6] Zhang, J. et al. (2011). Data-driven intelligent transportation systems: A survey. IEEE Trans. Intell. Transp. Syst., 12(4), 1624–1639.

[7] Herrera, J. C., Work, D. B., Herring, R., Ban, X. J., Jacobson, Q., & Bayen, A. M. (2010). Evaluation of traffic data obtained via GPS-enabled mobile phones: The Mobile Century field experiment. *Transportation Research Part C: Emerging Technologies*, *18*(4), 568-583.

[8] Hoseinzadeh, N., Arvin, R., Khattak, A. J., & Han, L. D. (2020). Integrating safety and mobility for pathfinding using big data generated by connected vehicles. *Journal of Intelligent Transportation Systems*, *24*(4), 404-420.

[9] CENTER, S. T. (2016). BIG DATA GENERATED BY CONNECTED AND AUTOMATED VEHICLES FOR SAFETY MONITORING, ASSESSMENT AND IMPROVEMENT.

[10] Mohammad Nazar, A., Arvin, R., & Khattak, A. J. (2021). Classifying travelers' driving style using basic safety messages generated by connected vehicles: Application of unsupervised machine learning. *Transportation research part C: emerging technologies*, *122*, 102917.

[11] Maryam Mousavi, S., Lord, D., Dadashova, B., & Reza Mousavi, S. (2020, August). Can autonomous vehicles enhance traffic safety at unsignalized intersections? In *International Conference on Transportation and Development 2020* (pp. 194-206). Reston, VA: American Society of Civil Engineers.

[12] Mousavi, S. M., Osman, O. A., Lord, D., Dixon, K. K., & Dadashova, B. (2021). Investigating the safety and operational benefits of mixed traffic environments with different automated vehicle market penetration rates in the proximity of a driveway on an urban arterial. *Accident Analysis & Prevention*, *152*, 105982.

[13] Hasoon, S. T., & Mahdi, M. A. (2017). A Developed Realistic Urban Road Traffic in Erbil City Using Bi-directionally Coupled Simulations. *Qalaai Zanist Journal*, *2*(2), 27-34.

[14] Azad, M., Hoseinzadeh, N., Brakewood, C., Cherry, C. R., & Han, L. D. (2019). A literature review on fully autonomous buses. *Transportation Research Board 98th Annual MeetingTransportation Research Board*, (19-05492).

[15] Mahdinia, I., Mohammadnazar, A., Arvin, R., & Khattak, A. J. (2021). Integration of automated vehicles in mixed traffic: Evaluating changes in performance of following human-driven vehicles. *Accident Analysis & Prevention*, *152*, 106006.

[16] Mahdinia, I., Arvin, R., Khattak, A. J., & Ghiasi, A. (2020). Safety, energy, and emissions impacts of adaptive cruise control and cooperative adaptive cruise control. *Transportation Research Record*, *2674*(6), 253-267.

[17] Chatzimilioudis, G., Konstantinidis, A., Laoudias, C., & Zeinalipour-Yazti, D. (2012). Crowdsourcing with smartphones. *IEEE Internet Computing*, *16*(5), 36-44.

[18] Tian, Y., Zhang, K., Li, J., Lin, X., & Yang, B. (2018). LSTM-based traffic flow prediction with missing data. *Neurocomputing*, *318*, 297-305.

[19] Ryu, U., Wang, J., Kim, T., Kwak, S., & Juhyok, U. (2018). Construction of traffic state vector using mutual information for short-term traffic flow prediction. *Transportation Research Part C: Emerging Technologies*, *96*, 55-71.

[20] Deng, S., Jia, S., & Chen, J. (2019). Exploring spatial–temporal relations via deep convolutional neural networks for traffic flow prediction with incomplete data. *Applied Soft Computing*, *78*, 712-721.

[21] Sekuła, P., Marković, N., Vander Laan, Z., & Sadabadi, K. F. (2018). Estimating historical hourly traffic volumes via machine learning and vehicle probe data: A Maryland case study. *Transportation Research Part C: Emerging Technologies*, *97*, 147-158.

[22] Altintasi, O., Tuydes-Yaman, H., & Tuncay, K. (2017). Detection of urban traffic patterns from Floating Car Data (FCD). *Transportation research procedia*, *22*, 382-391.

[23] Jolovic, D., Stevanovic, A., Sajjadi, S., & Martin, P. T. (2016). Assessment of level-of-service for freeway segments using HCM and microsimulation methods. *Transportation Research Procedia*, *15*, 403-416.

[24] Celikoglu, H. B., & Silgu, M. A. (2016). Extension of traffic flow pattern dynamic classification by a macroscopic model using multivariate clustering. *Transportation Science*, *50*(3), 966-981.

[25] Aljamal, M. A., Abdelghaffar, H. M., & Rakha, H. A. (2019). Developing a neural–Kalman filtering approach for estimating traffic stream density using probe vehicle data. *Sensors*, *19*(19), 4325.

[26] Wassantachat, T., Li, Z., Chen, J., Wang, Y., & Tan, E. (2009, September). Traffic density estimation with on-line SVM classifier. In *2009 Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance* (pp. 13-18). IEEE.

[27] Kurniawan, F., Sajati, H., & Dinaryanto, O. (2017). Image processing technique for traffic density estimation. *International Journal of Engineering and Technology*, *9*(2), 1496-1503.

[28] Kittelson, W., & Vandehey, M. (2013). *Incorporation of travel time reliability into the HCM* (No. SHRP 2 Reliability Project L08).

[29] SINGH, V., GORE, N., CHEPURI, A., ARKATKAR, S., JOSHI, G., & PULUGURTHA, S. (2019). Examining travel time variability and reliability on an urban arterial road using wi-fi detections-A case study. *Journal of the Eastern Asia Society for Transportation Studies*, *13*, 2390-2411.

[30] Kodupuganti, S. R., & Pulugurtha, S. S. (2019). Link-level travel time measures-based level of service thresholds by the posted speed limit. *Transportation research interdisciplinary perspectives*, *3*, 100068.

[31] Pulugurtha, S. S., & Imran, M. S. (2020). Modeling basic freeway section level-of-service based on travel time and reliability. *Case Studies on Transport Policy*, *8*(1), 127-134.

[32] McHugh, C., Coleman, S., & Kerr, D. (2021). Technical indicators for energy market trading. *Machine Learning with Applications*, *6*, 100182.

https://doi.org/10.25130/tjps.v28i3.1428

[33] Alhashel, B. S.; Almudhaf, F. W. and Hansz, J. A. (2018). Can technical analysis generate superior returns in securitized property markets? Evidence from east asia markets. Pacific-Basin Finance Journal, 47, 92–108.

[34] Romero-del-Castillo, J.A.; Mendoza-Hurtado, M.; Ortiz-Boyer, D. and García-Pedrajas, N. (2022). Local-based k values for multi-label k-nearest neighbor's rule. Engineering Applications of Artificial Intelligence, 116: 105487.

[35] England, H. (2015). 'Highways England Network Journey Time and Traffic Flow Data-Webtris'. *Open Government License v3*. [Access May 2022].

[36] Tanaka-Yamawaki, M., & Tokuoka, S. (2007). Adaptive use of technical indicators for the prediction of intra-day stock prices. Statistical Mechanics and Its Applications, 383, 125–133. http://dx.doi.org/10.1016/j.physa.2007.04.126.

[37] Pring, M. J. (2002). *Technical analysis explained: The successful investor's guide to spotting investment trends and turning points*. McGraw-Hill Professional.

.

[38] Ayala, J., García-Torres, M., Noguera, J. L. V., Gómez-Vela, F., & Divina, F. (2021). Technical analysis strategy optimization using a machine learning approach in stock market indices. *Knowledge-Based Systems*, *225*, 107119.

[39] Weng, B.; Ahmed, M. A.; and Megahed, F. M. (2017). Stock market one-day ahead movement prediction using disparate data sources. Expert Systems with Applications, 79: 153–163.

[40] Kim, M. K.; Kim, Y. S. and Srebric, J. (2020). Predictions of electricity consumption in a campus building using occupant rates and weather elements with sensitivity analysis: Artificial neural network vs. linear regression. Sustainable Cities and Society, 62: 102385.

[41] Saltelli, A. (2002). Sensitivity analysis for importance assessment. Risk Analysis, 22(3), 579–590.

[42] Pedregosa, F. et al. (2011). Scikit-learn: Machine Learning in Python. J. Mach. Learn. Res., 12: 2825–2830.

[43] T. M. Units. (2018). "National Traffic Information Service DATEX II Service".