*sustainability*

MDPI

*Article*

# Improving Traffic-Flow Prediction Using Proximity to Urban Features and Public Space

**Rawan Rajha [1], Shino Shiode [1] and Narushige Shiode [2,*]**

[1] Department of Geography, Birkbeck, University of London, London WC1E 7HX, UK; rrajha01@student.bbk.ac.uk (R.R.); s.shiode@bbk.ac.uk (S.S.)

[2] Department of Geography, Geology and the Environment, Kingston University, Kingston upon Thames KT1 2EE, UK

* Correspondence: n.shiode@kingston.ac.uk

**Abstract:** Accurate traffic prediction and planning help alleviate congestion and facilitate sustainable traffic management through short-term traffic controls and long-term infrastructure design. While recent uptake on Machine Learning (ML) approaches helps refine our ability to predict the traffic flow, proximity to landmarks and public spaces are often overlooked, thus undermining the impact of location-specific traffic patterns. Using traffic-flow estimates from London, this study incorporates the proximity to urban features approximated with Kernel Density Estimation (KDE) and compares the performance of models with and without such features. They are also tested using classic spatial/non-spatial regression models and ML-based regression models. Results suggest that adding urban features considerably improves the performance of the ML models (Fine tree yielding $R^2$ = 0.94, $RMSE$ = 0.129, and $MAE$ = 0.069), which compares favourably against the best performing non-ML model (the spatial error model returning $R^2$ = 0.448, $RMSE$ = 0.358, and $MAE$ = 0.280). Sensitivity of the KDE is tested across different bandwidths for including urban features. The ML classification approach was also applied for estimating the traffic density and achieved high accuracy, with Fine KNN achieving 98.7%. They offer a robust framework for accurate traffic projection at specific locations, thus enabling road infrastructure designs that cater to the specific needs of the local situations.

**Keywords:** machine learning; spatial modelling; traffic flow; traffic prediction



Academic Editor: Xinqiang Chen and Bin Ji

Received: 29 August 2024
Revised: 15 December 2024
Accepted: 23 December 2024
Published: 26 December 2024

**Citation:** Rajha, R.; Shiode, S.; Shiode, N. Improving Traffic-Flow Prediction Using Proximity to Urban Features and Public Space. *Sustainability* **2025**, *17*, 68. https://doi.org/10.3390/su17010068



## 1. Introduction

Rapid urban growth has made traffic-flow prediction indispensable for the economic and sustainable development of urban space. Making an accurate prediction of traffic flows can help alleviate traffic congestion [1]. Indeed, traffic-flow prediction can feed the road users with live traffic information, prompt the traffic controllers to enforce short-term solutions [2], whilst also enabling planners to make informed decisions on the operation and management of the road network (e.g., policy-led solutions for speed control, carbon emission, and operation of autonomous vehicles) [3]. They could yield positive impact on public safety, economy, as well as the environment and their quality of life. The World Health Organization noted that road traffic fatality was on the rise with 1.35 million deaths being recorded in 2016, and that traffic prediction would not only help reduce congestion but also facilitate safer and more cost-effective travels [4].

Conventional statistical models such as ordinary least squares (OLS) regression, spatial regression models, Autoregressive Integrated Moving (ARIMA), and Seasonal ARIMA offer a clear and robust framework for making such predictions. Yet, they are limited in their scope to process large, complex, and non-linear data structure. This is reflected in their choice of variables and the moderate accuracy of their prediction [5]. To overcome such limitations and to improve on their performance, recent studies have increasingly utilised Machine Learning (ML) approaches for predicting the traffic flows. Their capacity to process large non-linear data generally improved the performance of the prediction models [2,6]. However, the choice of the data and the variables used in the existing range of ML-based models would benefit from further scrutiny, especially those pertaining to the spatial features [7]. Traffic-flow prediction involves several factors, including traffic volume, vehicle speed, vehicle counting, road characteristics, as well as seasonal and weather information [8–10]. Some of these factors show spatial dependency (e.g., road characteristics and the land use tend to be similar within small, confined areas) and, yet, their traits are not always incorporated into the prediction model.

Against this background, this study develops a series of prediction models that incorporates the spatial proximity to the relevant features into ML methods and applies it to traffic estimate data from the Great London area. Specifically, we will (1) identify suitable predictors that are underused in the existing traffic prediction models, (2) incorporate them into ML prediction models and evaluate their performance, (3) compare the performance of the ML approach and the traditional statistical approaches by incorporating these predictors, and (4) conduct sensitivity analysis using the kernel density bandwidth. The focus of our study lies on the development of offline models for traffic-flow prediction and traffic density estimation, accounting for the urban features that are considered to affect the local traffic patterns. A systematic evaluation is offered through the application of 21 ML and non-ML methods against different combinations of variables that reflect proximity to these urban features.

## 2. Literature Review

### *2.1. Traffic Features and Traffic Indicators*

Traffic-flow prediction offers the short-term benefit of informing the road users whilst also providing the long-term benefit of helping build more efficient, sustainable, and liveable cities. The performance of the traffic-flow prediction largely depends on the modelling framework, the predictors, as well as the quality and the granularity of the data used. Prevailing predictors are often categorised into traffic features (e.g., road geometry, angles at the junction, number of lanes) and indicators (e.g., transport facilities, volume, and speed), and they both tend to have a strong impact on the model performance [11–14].

Some studies have also incorporated spatial features into their model, and these include locations of POIs (Point of Interests) and their impact on traffic accidents, specific land use (e.g., residential facilities and offices), and other landmarks [7,11,15–17].

Other studies noted the importance of understanding the impact of spatial features on traffic characteristics, taking a holistic approach to account for multiple features, whilst also observing the ways in which adjacent roads are connected, both of which are reported to improve the prediction performance [18,19]. For instance, Zeng et al. (2022) [11] examined how the presence of specific POIs would affect the traffic flows and heighten the risks of traffic accidents. The study suggests that there is both temporal and spatial impact of POIs on the local traffic flow. Other studies suggest the relevance of bus stops and how that can impact the traffic flow, especially considering their locations on busy roads and intersections. Similarly, train stations are considered to have a strong impact on the traffic flow. Hospital locations, schools, and other landmark that attract a large number of people are also prone to affecting the local traffic flow, owing to their influence on road network capacity, and thus often leading to bottlenecks or congestion points.

POIs of recreational purposes (e.g., large urban parks) and tourist attractions also tend to form a hub for congestion hotspots, attracting more traffic in the area. They serve as external factors on traffic-flow prediction, including tourist places. Finally, urban parks and landscape areas (e.g., Hyde Park and Regent Park) are usually located near busy roads or intersections. As they may restrict vehicles from crossing the area and thus force road users to find alternative routes, that can also lead to increased congestion.

Traffic indicators, on the other hand, provide information on the state and the condition of the road network that could affect the performance of traffic on a road network. They are classified into two categories: traffic state and traffic-flow parameters [20]. Traffic state focuses on the impact of traffic variables on road segments, using partially observable traffic data, while traffic-flow parameters (e.g., average speed and flow rate) are used for deciphering traffic-flow characteristics and predicting changes in their behaviour [21]. An increasing number of traffic-flow prediction models utilise traffic state and parameters [22], and their performance is based on the quality and the quantity of the available data [23].

### *2.2. Prediction Models: ML and Non-ML Methods*

In terms of the model architecture, existing traffic-flow models can be classified into (1) the conventional statistical methods and (2) the ML and deep learning (DL) approaches. Use of ML and DL techniques for traffic prediction has been prominent in recent years, supported by the increasingly available traffic data and computational resources [24]. They allow us to tackle complex traffic prediction problems more efficiently than statistical methods can [25].

Statistical models, such as OLS, K-Nearest Neighbour (KNN), Historical Average, and Vector Autoregressive are often used for short-term traffic prediction owing to their simple and clear computational structure as well as their capacity to facilitate robust theoretical interpretation [26–31]. Parametric techniques such as ARIMA and Kalman filtering are also used for predicting the temporal changes in the traffic flow [32,33]. These methods are suitable for small and confined datasets, but they are not designed to handle a complex or dynamic system that exhibits non-linear behaviour [34].

ML models can largely overcome the inherent limitations of the statistic models in dealing with complex datasets. Specifically, they can analyse and process large and complex datasets with their ability to generalise and learn complex relationships and, thereby, extract meaningful patterns and associations among the attributes within the data [7,35]. They can also adapt to the dynamic conditions of the traffic network [36].

ML approaches for traffic-flow prediction can be categorised into two sub-categories: classification and regression. Classification helps predict a categorical variable such as the state of traffic flow or the level of congestion. Some traffic-focused classifications include the application of KNN [27] and random forest for predicting the level of traffic flow [37]. They deal with discrete classes and tend to be computationally efficient.

Regression, on the other hand, uses a set of predictors to estimate or forecast the dependent variables that take a continuous range (e.g., traffic-flow volume, traffic speed). Examples include the use of Ensemble tree ML regression to estimate the impact of road accidents on traffic flow [38], and ML regression for predicting traffic flow with the effect of data sharing across the Internet of Vehicles environment [39].

Both classification and regression algorithms use training data with known predictors or input features and the corresponding target values to unravel the underlying patterns and relationships. Indeed, ML models such as Support Vector Machine (SVM) and Decision trees can be used for both regression and classification tasks. Table 1 shows a list of ML models and their suitability for these tasks.

**Table 1.** Machine learning methods and their description, with relevant references.

| Model | Description |
|---|---|
| Support Vector Machine (SVM) [35] | Used as both classifier and regressor. It classifies data into different classes and estimates the optimal boundary values for classes to predict the value of a target variable based on input variables. SVMs perform well when there is a clear separation between the classes and the number of input variables is high. The model is also known for its ability to hold high-dimensional data and can deal with multitude of missing data. However, SVM is complex, resource-intensive, and is sensitive to the configuration parameters (e.g., regularisation parameter and kernel). |
| Neural Network (NN) [35,40] | Used as both classifier and regressor. It works through a network of multiple interconnected layers to generate prediction. NN can learn complex non-linear relationships between the input variables and outputs. However, NN requires a large and good-quality training dataset, and the use of small datasets could result in overfitting. |
| Decision trees [41,42] | Used as both classifier and regressor. Arguably the most common ML methods with a range of applications. It works through a tree-like structure comprising nodes and branches to develop a prediction model. Decision trees are easy to interpret and can provide valuable insights into the most important input variables for predicting target variables. They provide advantages in handling mixed data (categorical and numerical) and are less susceptible to the impact of outliers than other ML algorithms are. However, they have challenges in coping with large datasets and are prone to overfitting. |
| K-Nearest Neighbour (KNN) [20,28] | Used as both classifier and regressor. KNN is based on searching the historical database for the k events that are closest to the current traffic situation, which are subsequently averaged or weighted by their distance to the current situation. KNN is a smoothing method that uses more information compared to the present by considering multiple close matches. KNN, however, is resource-intensive, as it searches through the historical database and is sensitive to the change in k values. |
| Gaussian Process Regression (GPR) [43,44] | Used as a regressor only but is widely used for traffic-flow prediction. These models leverage the flexibility and the stochastic nature of Gaussian processes to capture the complex and non-linear relationships between traffic variables. GPR is sensitive to the configuration parameters (Kernel and Hyperparameters), and takes a complex high-dimensional form. |

### *2.3. Deep Learning Methods*

Alongside the ML models, DL methods are increasingly used for traffic-flow prediction, especially in terms of improving the model performance when handling data from different sources and for offering real-time predictions. While DL offers higher precision on real-time applications, it is more complex than ML methods. The most common and widely used DL models in the traffic prediction context include the following:

(1) Convolutional Neural Networks (CNNs) utilise a convolution matrix to capture the long-range spatial dependence of traffic flows. CNN is effective in predicting traffic flows, especially in case of a complex traffic pattern with multiple factors affecting the traffic flow [2,34].

(2) Recurrent Neural Networks (RNNs) are widely used for time-series data processing. They use Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRUs) as

RNN variants to facilitate real-time traffic prediction. RNN can capture temporal dependencies in sequential data and predict future traffic flow [45,46].

(3) Graph Convolutional Networks (GCNs) are specifically designed for processing graph-structured data. The graphs are used for reflecting the degree of association between features and the traffic flow, and are utilised for real-time operations [47,48].

(4) Deep Neural Network (DNN) can learn complex patterns from large datasets. It connects the input layer to one or more hidden layers, and the output layer generates the predicted traffic flow [49–51].

While DL models are known to deliver improved performance on traffic-flow prediction, they also suffer from potential overfitting and limited interpretability. As these traits do not align well with the scope of our study on evaluating the impact of POIs and urban facilities on traffic-flow prediction, we will focus on the mainstream ML methods and non-ML methods.

### 2.4. Gap in the Literature

While ML and DL methods are increasingly prevalent in traffic-flow modelling, the following challenges remain:

(1) Urban features such as POIs and public space are underused as predictors in the context of traffic-flow prediction. As the choice of predictors could significantly affect the accuracy of the prediction, selecting the right set of features is critical for developing an accurate and robust model. Nevertheless, existing studies have focused mainly on refining their algorithms with less attention to incorporating the relevant spatial features [11,17].

(2) The few studies that account for urban features in the vicinity of each data point have used KDE and adopted a predetermined value for its bandwidth. This value is usually fixed a priori with no sensitivity analysis of the bandwidth. For instance, Zeng et al. (2022) [11] used KDE in conjunction with CNN to incorporate the spatial features, while Jia et al. (2018) [15] applied KDE for OLS and spatial regression models for traffic crashes. Neither study evaluates the impact of KDE bandwidth on the respective ML model.

To offer a systematic evaluation on the validity and the sensitivity of the KDE for traffic-flow modelling, a comprehensive framework is needed so that it can combine all relevant spatial features and traffic indicators to improve prediction accuracy. There is also a need for comparative evaluations of various ML models to identify the most accurate model for predicting the traffic flow. Examining the sensitivity of KDE by changing its bandwidth would allow us to empirically derive the optimal bandwidth for traffic prediction, although it may vary between different countries and cities.

## 3. Methodology

### 3.1. Selecting the ML and the Non-ML Methods

To develop a robust and accurate model framework for traffic prediction, a total of 21 different models are constructed, consisting of 3 non-ML methods (non-spatial, OLS regression; spatial lag model; and spatial error model) and 18 ML methods from across 5 families of ML approaches. They consist of the traditional mainstream ML models of tree-based, Ensemble, SVM, GPR, and NN (as shown in Table 1) as well as their variants: namely, Decision trees (fine, medium, and coarse), Decision tree ensembles (bagging, boosting), Support vector machines (cubic, fine Gaussian, and medium Gaussian), Gaussian process regressions (squared, exponential, and Matern covariogram), Kernels (rational quadratic, Support vector machine, and least-squares), and Neural Networks (narrow, wide, bilayered, and trilayered)). The choice of these methods was made on the basis of their collective ability to offer a comprehensive evaluation of the performance of each combination of the variables (which we hereafter refer to as scenarios; please refer to Section 3.2 for the five scenarios used in our study). The selection of models was also designed to accommodate a range of possible associations between the variables; namely linear, non-linear, and complex interactions.

Figure 1 shows the framework of our traffic-flow prediction modelling. It is designed to address the research gaps identified and carry out the following steps for traffic prediction:

(1) Introduce a combination of relevant urban and spatial features into the traffic-flow prediction model. The literature suggests that the urban features and the traffic characteristics used in our study are considered to affect the local traffic patterns. They are selected through either the review of the relevant literature (Table 2) [11,12,52–54] or the prevailing trend on traffic-flow modelling. Their sphere of influence is represented by means of a smoothing technique called Kernel Density Estimation (KDE).

(2) Develop non-ML regression models (namely, OLS and spatial regression models) using features identified in the literature review.

(3) Develop a framework of ML regression models for traffic-flow prediction. We conduct grid search and cross-validation to optimise the parameters such as tree depth, SVM kernel settings, and the size of the Neural Network layers, establishing a suitable balance between the model complexity and performance. Each ML model is trained using multiple parameter configurations, including combinations of widely recognised parameters.

(4) Assess the impact of utilising a combination of spatial and non-spatial features to improve traffic prediction models. All experiments are conducted on a standardised computational setup, with specifications noted for clarity and reproducibility. Two programming languages are used for training the models (python 3.7.6 and MATLAB R2022b).

(5) Identify the best performing ML methods for traffic-flow prediction and traffic density estimation. Also, compare their performance against those derived by OLS and spatial regression models.

(6) Conduct sensitivity analysis to identify the optimal kernel density bandwidth that can improve the accuracy of prediction. The impact of urban features is weighted with a predetermined KDE bandwidth of 1000 m. To determine the most relevant features and their sphere of influence, five experiments are conducted in this framework.

**Table 2.** Features to be used as predictors in the ML prediction models.

| Categories | Features | Description |
|---|---|---|
| Road characteristics | Road type | Each road type has different capacities and speed limits. For instance, high-capacity roads (e.g., major roads) are designed to hold large volumes of traffic moving at higher speeds, and they naturally have high traffic flow. Road types have a significant influence on the traffic [52]. |
| | Road environment | It represents the road nature, including the presence of a traffic island link, proximity to a junction, dual carriageway, single carriageway, and the presence of a slip road. The road environment (the number of lanes in particular) impacts traffic flow, as it is directly linked to the road capacity and the traffic volume [52]. |
| | Road shape/length | The geometrical features of roads, including the road length and their shape, are considered to affect the traffic flow [12]. Longer roads may introduce variations in the traffic flow, due to changes in road characteristics, intersections, access points, and other factors alongside their length. |

**Table 2.** *Cont.*

| Categories | Features | Description |
|---|---|---|
| Facilites | Bus stops | Bus stops can impact the traffic flow, considering their locations on busy roads and intersections, where buses may hold traffic whilst loading and unloading passengers. |
| | Train stations | Train stations have a strong impact on traffic flow [55]. In general, access to public transport nodes is considered to affect traffic-flow prediction (e.g., the multifaceted impacts on urban road systems) [53,54]. |
| Point of Interest | Hospitals | Hospital locations may also affect the traffic flow owing to their influence on road network capacity, leading to bottlenecks or congestion points [56]. |
| | Colleges | A set of other POIs are also used to evaluate the temporal and spatial impact of POIs on the traffic flow. Their impact is estimated by measuring the distance from the road surface and traffic sensors, so as to improve the prediction accuracy [11]. |
| Public space | Recreational space | POIs of recreational purposes may impact traffic-flow prediction. Recreational space can attract a high volume of vehicles during peak hours or special events. Increased traffic volume around these places can lead to localised congestion and longer travel times in the surrounding area. |
| | Tourism places | Tourist attractions often become congestion hotspots, as there is often an increase in the traffic volume as well as loading and unloading of the passengers around tourism destinations. They serve as external factors on traffic-flow prediction including tourist places, considered an attraction point for crowdedness in urban areas. |
| Landscape | | Landscape areas, such as Hyde Park and Regent Park, are usually located near busy roads or intersections, as they may restrict vehicles from crossing the area and thus force road users to find alternative routes, which can also lead to increased congestion. |

### *3.2. Traffic Prediction Using Different Combinations of Features*

Table 2 lists the urban features and road characteristics used in our traffic-flow prediction modelling. The specific rationale for adopting each variable is described under each feature entry. In essence, they are considered in the existing literature as having a non-trivial impact on the traffic flow and, by comparing the prediction outcomes based on the use of different combinations of these features, we will offer a systematic evaluation of how these categories of features can be used towards explaining and predicting the traffic flow around them.

To derive prediction from the ML methods, we use them as either a regressor (for the traffic flow) or a classifier (for the traffic density) so that we can improve the overall performance of predicting the traffic situation. The predictors will be determined by adding different combinations of features to the prediction model, as listed below (Table 3):

- Scenario 1 uses the basic traffic indicators (average traffic density, hour, road segment, and days) to establish a baseline prediction model.
- Scenario 2 uses the basic traffic indicators and the road characteristics (road type, road nature, and road length) to assess the impact of road design on traffic flow.
- Scenario 3 uses the basic traffic indicators and spatial features (POIs such as hospitals and colleges, as well as public nodes and public transport facilities).
- Scenario 4 uses the basic traffic indicators, road characteristics, and two of the POIs (hospitals and colleges).
- Scenario 5 uses all feature categories, namely the basic traffic indicators, road characteristics, as well as all spatial features.

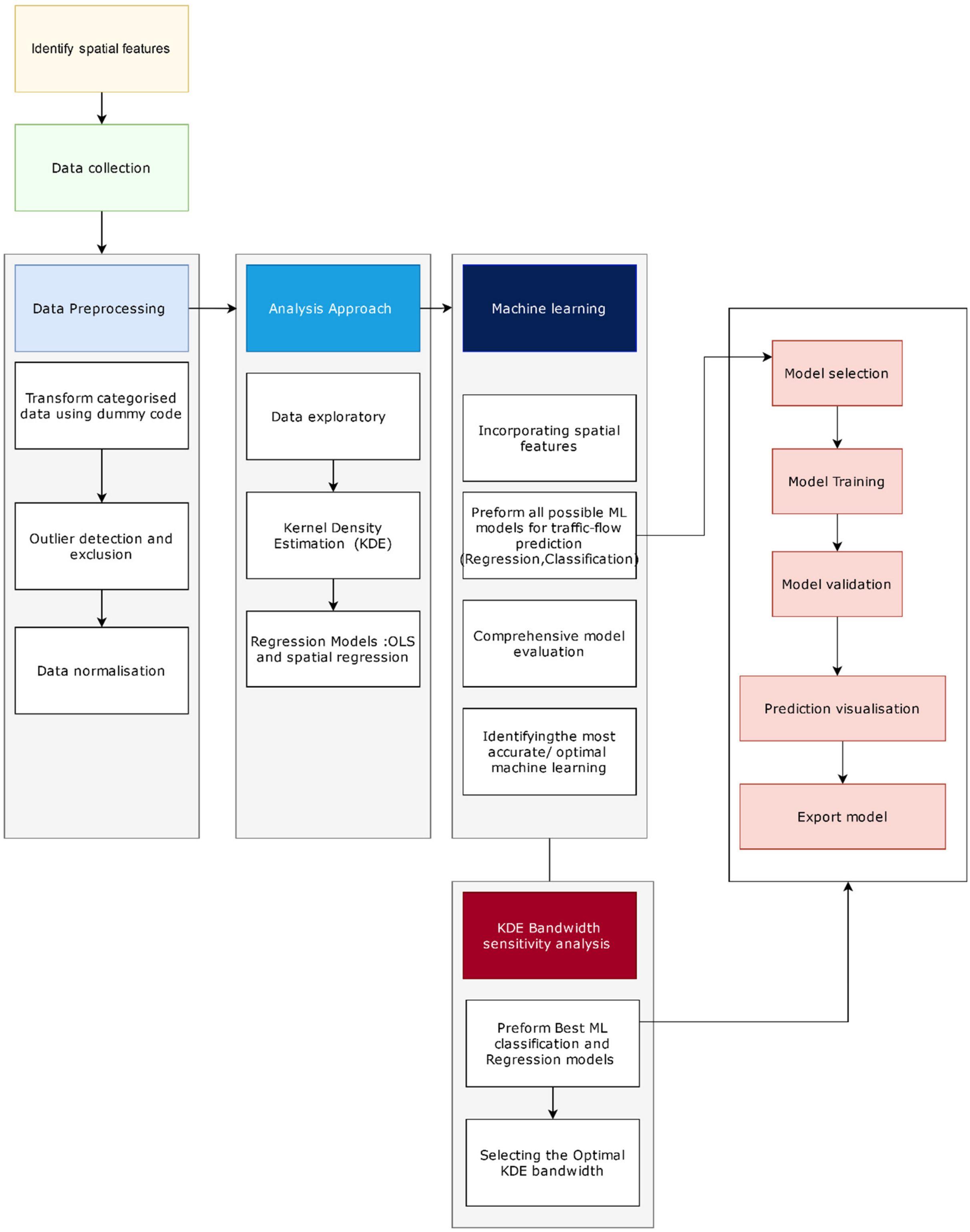


**Figure 1.** Flow of research for traffic modelling.

The difference in the combinations of the predictors between these scenarios will enable us to assess the performance of a particular type of predictors in refining the prediction.

**Table 3.** The five scenarios' combinations of predictors for the ML model.

| | **Traffic Indicators** | **Road Characteristics** | **Spatial Features** | | | |
|---|---|---|---|---|---|---|
| | Traffic Density by Hours, Days, Road Segments | Road Type, Road Nature, Road Length | POIs (Urban Facilities) | | | Public Transport Facilities |
| | | | Hospitals | Colleges | Public Nodes * | Bus Stops |
| Scenario 1 | √ | | | | | |
| Scenario 2 | √ | √ | | | | |
| Scenario 3 | √ | | √ | √ | √ | √ |
| Scenario 4 | √ | √ | √ | √ | | |
| Scenario 5 | √ | √ | √ | √ | √ | √ |

* Landscape areas, tourist places, and recreational places.

### *3.3. Traffic Data*

Our study uses the traffic-flow data from London, UK. They are a set of annual road traffic estimates provided by the UK Road Traffic, Department for Transport (roadtraffic.dft.gov.uk [accessed on 24 April 2023]). It comprises street-level road traffic estimates, which are derived through a combination of automatic traffic counts from around 300 locations and approximately 8000 manual traffic counts carried 1000 out annually for the Department for Transport's road traffic statistics (conducted by trained enumerators from 7 a.m. to 7 p.m. on weekdays between March and October). As it takes the form of annual street-level road traffic estimates with in-house preprocessing, no records were shown as missing their fields or deemed incomplete, i.e., no data was truncated or eliminated for incomplete entries, nor was any anomaly noted. This dataset caters the scope of our study well, as we aim to develop offline models to evaluate the impact of urban facilities on predicting traffic flows. However, any follow-up studies intended for developing a real-time prediction may require training of the models with disaggregate real-world data that may be missing some fields and/or contain anomalies. The impact of such localised and short-term anomalies will also require a separate investigation.

## 4. Analysis

### *4.1. Machine Learning Model: Regression*

As described earlier, a total of 18 supervised ML methods were used across five families for predicting the traffic flow, and they were applied to all five scenarios (Table 3). The results are shows in Table 4.

They confirm that the ML methods vastly outperform the classic regression models. In particular, the validation metrics of the OLS regression model, spatial error, and spatial lag were ($R^2$ = 0.30, *RMSE* = 0.397, and *MAE* = 0.288), ($R^2$ = 0.448, *RMSE* = 0.358, and *MAE* = 0.280), and ($R^2$ = 0.438, *RMSE* = 0.388, and *MAE* = 0.282), respectively; whereas ML (Fine tree) applied to the same set of predictors returns ($R^2$ = 0.94, *RMSE* = 0.129, and *MAE* = 0.069). These results clearly illustrate the efficiency of ML methods in predicting the traffic flow.

The ML methods generally perform much better when the spatial features are added as predictors. For instance, using the traffic indicators alone (Scenario 1) only returned a modest performance, whereas combining the traffic indicators with the road characteristics

(Scenario 2) saw a considerable improvement in their performance. Similarly, the use of the spatial features (POIs and bus stops) significantly improved the outcomes, particularly the GPR and Bagged tree models, while the road characteristics in Scenario 2 helped improve the performance for SVM, Bilayered NN, and narrow NN. Between Scenarios 2 and 3, the spatial features had a higher impact than road characteristics on the final models.

**Table 4.** Performance of ML traffic-flow prediction against Scenarios 1–5.

| | **Scenario 1** | | | **Scenario 2** | | | **Scenario 3** | | | **Scenario 4** | | | **Scenario 5** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *RMSE* | $R^2$ | *MAE* | *RMSE* | $R^2$ | *MAE* | *RMSE* | $R^2$ | *MAE* | *RMSE* | $R^2$ | *MAE* | *RMSE* | $R^2$ | *MAE* |
| Decision tree (fine) | 0.311 | 0.57 | 0.211 | 0.137 | 0.91 | 0.074 | 0.140 | 0.90 | 0.072 | 0.133 | 0.92 | 0.070 | 0.129 | 0.94 | 0.069 |
| Decision tree (med) | 0.303 | 0.59 | 0.211 | 0.157 | 0.88 | 0.090 | 0.161 | 0.89 | 0.089 | 0.163 | 0.89 | 0.093 | 0.145 | 0.91 | 0.082 |
| Decision tree (coarse) | 0.321 | 0.54 | 0.230 | 0.234 | 0.76 | 0.150 | 0.233 | 0.76 | 0.145 | 0.232 | 0.76 | 0.149 | 0.190 | 0.84 | 0.133 |
| SVM Cubic | 0.395 | 0.31 | 0.258 | 0.364 | 0.42 | 0.231 | 0.375 | 0.38 | 0.242 | 0.353 | 0.45 | 0.222 | 0.208 | 0.68 | 0.238 |
| SVM Fine Gaus. | 0.384 | 0.38 | 0.245 | 0.277 | 0.66 | 0.161 | 0.267 | 0.69 | 0.154 | 0.231 | 0.77 | 0.131 | 0.178 | 0.84 | 0.202 |
| SVM Med.Gaus. | 0.388 | 0.37 | 0.245 | 0.348 | 0.47 | 0.217 | 0.356 | 0.45 | 0.224 | 0.341 | 0.49 | 0.208 | 0.197 | 0.83 | 0.106 |
| Ensemble (Boosted) | 0.333 | 0.51 | 0.244 | 0.289 | 0.63 | 0.211 | 0.291 | 0.63 | 0.212 | 0.290 | 0.63 | 0.211 | 0.263 | 0.69 | 0.194 |
| Ensemble (Bagged ) | 0.322 | 0.54 | 0.231 | 0.226 | 0.78 | 0.158 | 0.145 | 0.91 | 0.083 | 0.166 | 0.88 | 0.108 | 0.130 | 0.92 | 0.074 |
| Squared GPR | 0.358 | 0.43 | 0.257 | 0.255 | 0.72 | 0.172 | 0.236 | 0.76 | 0.152 | 0.192 | 0.84 | 0.121 | 0.148 | 0.90 | 0.195 |
| Matern GPR | 0.354 | 0.45 | 0.255 | 0.242 | 0.74 | 0.160 | 0.222 | 0.79 | 0.139 | 0.187 | 0.85 | 0.117 | 0.130 | 0.90 | 0.075 |
| Exp. GPR | 0.340 | 0.49 | 0.232 | 0.232 | 0.76 | 0.145 | 0.236 | 0.76 | 0.145 | 0.198 | 0.83 | 0.121 | 0.148 | 0.87 | 0.084 |
| Rational Quad. | 0.338 | 0.50 | 0.237 | 0.237 | 0.75 | 0.152 | 0.227 | 0.77 | 0.143 | 0.184 | 0.85 | 0.112 | 0.149 | 0.91 | 0.084 |
| Narrow NN | 0.351 | 0.46 | 0.252 | 0.321 | 0.55 | 0.228 | 0.329 | 0.53 | 0.237 | 0.315 | 0.56 | 0.226 | 0.272 | 0.76 | 0.212 |
| Wide NN | 0.337 | 0.50 | 0.242 | 0.278 | 0.66 | 0.198 | 0.212 | 0.80 | 0.141 | 0.166 | 0.87 | 0.110 | 0.129 | 0.92 | 0.089 |
| Bilayered NN | 0.342 | 0.48 | 0.246 | 0.182 | 0.85 | 0.124 | 0.294 | 0.62 | 0.207 | 0.272 | 0.67 | 0.193 | 0.238 | 0.76 | 0.175 |
| Trilayered NN | 0.332 | 0.51 | 0.238 | 0.268 | 0.69 | 0.189 | 0.264 | 0.70 | 0.185 | 0.237 | 0.75 | 0.167 | 0.220 | 0.77 | 0.160 |
| SVM Kernel | 0.401 | 0.29 | 0.261 | 0.244 | 0.44 | 0.171 | 0.387 | 0.34 | 0.249 | 0.359 | 0.43 | 0.227 | 0.345 | 0.48 | 0.217 |
| LS Kernel | 0.384 | 0.35 | 0.279 | 0.364 | 0.42 | 0.229 | 0.375 | 0.39 | 0.274 | 0.355 | 0.44 | 0.255 | 0.340 | 0.49 | 0.245 |

Scenario 4 highlights the importance of using specific POIs, namely hospitals and colleges. Finally, Scenario 5 confirms that utilizing all features improves the overall model performance, with the Fine tree model returning the best performance with validation metrics at $R^2 = 0.94$, $RMSE = 0.129$, and $MAE = 0.069$, followed by Bagged tree and Wide NN models with validation metrics ($R^2 = 0.92$, $RMSE = 0.130$, and $MAE = 0.074$) and ($R^2 = 0.92$, $RMSE = 0.129$, and $MAE = 0.089$). Interestingly, the results from the Fine tree model are consistently high in performance across Scenarios 2–5, whereas other models are affected by the types and the combination of indicators. Nevertheless, the spatial features generally had the highest impact on improving the model performance between feature categories, which illustrates the importance of spatial features and analysis within traffic-flow prediction.

Figure 2 illustrates a summary of the highest $R^2$ values achieved by each ML family across five scenarios. While it confirms the consistently low performance of Scenario 1 (i.e., traffic characteristics alone), the difference between the remaining four scenarios varies from one ML family to another, and Scenario 5 facilitates the highest overall performance.

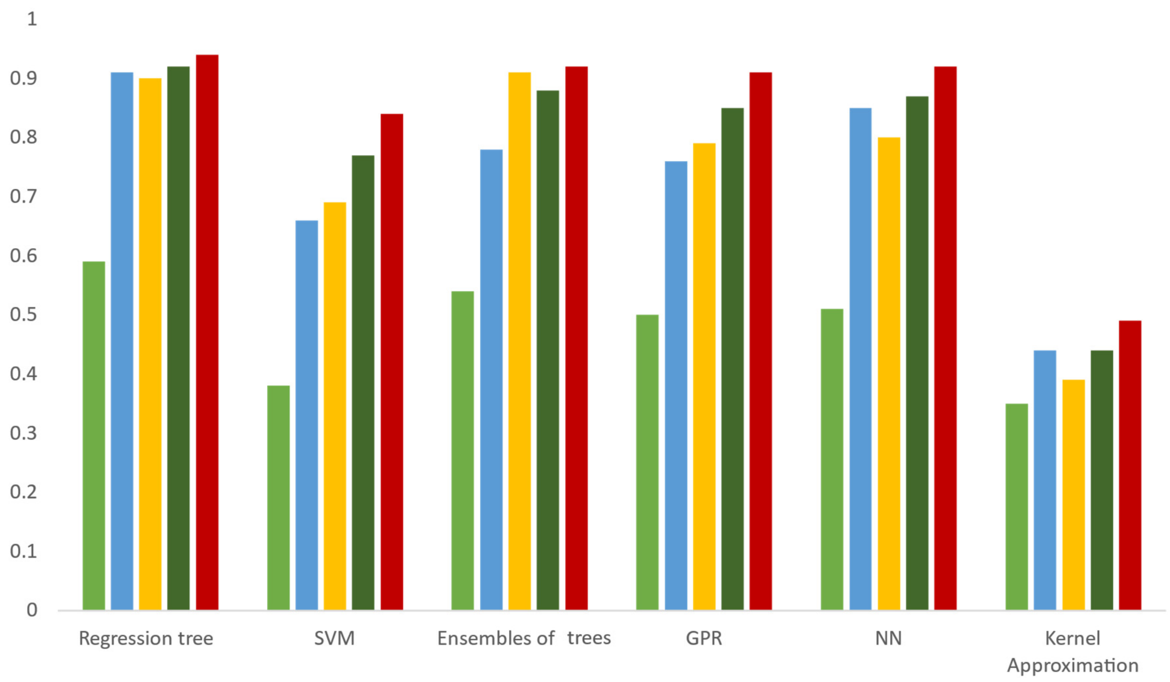


**Figure 2.** Highest $R^2$ values achieved by each ML family under the five scenarios.

Figure 3 shows the aggregate performance of all ML methods combined for Scenarios 1, 2, and 5. Scenarios 3 and 4 are omitted, as Scenario 5 represents their composite and yields better overall prediction performance than any ML methods. As expected, Scenario 5 returned the best overall performance compared to other scenarios. All families of ML methods achieved $R^2$ values above 0.90, except for the Kernel Approximation family which returned low performance across all five scenarios.

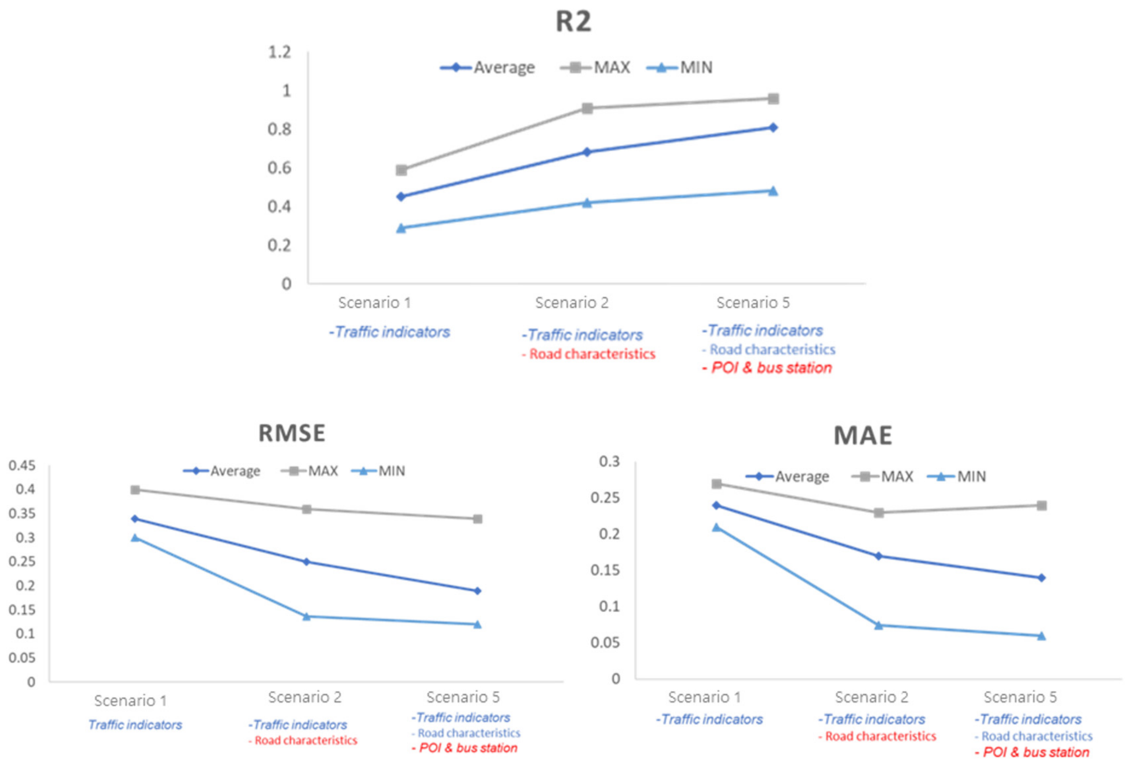


**Figure 3.** Validation metrics of the highest-performing ML methods across 5 scenarios.

Figure 4 shows the discrepancy between the expected and the observed values for Fine tree in Scenario 5, i.e., the best-performing ML method in Scenario 5. The fluctuation suggests that the predictive power is not consistent across the different values of observation index but returns an overall prediction that is highly accurate.

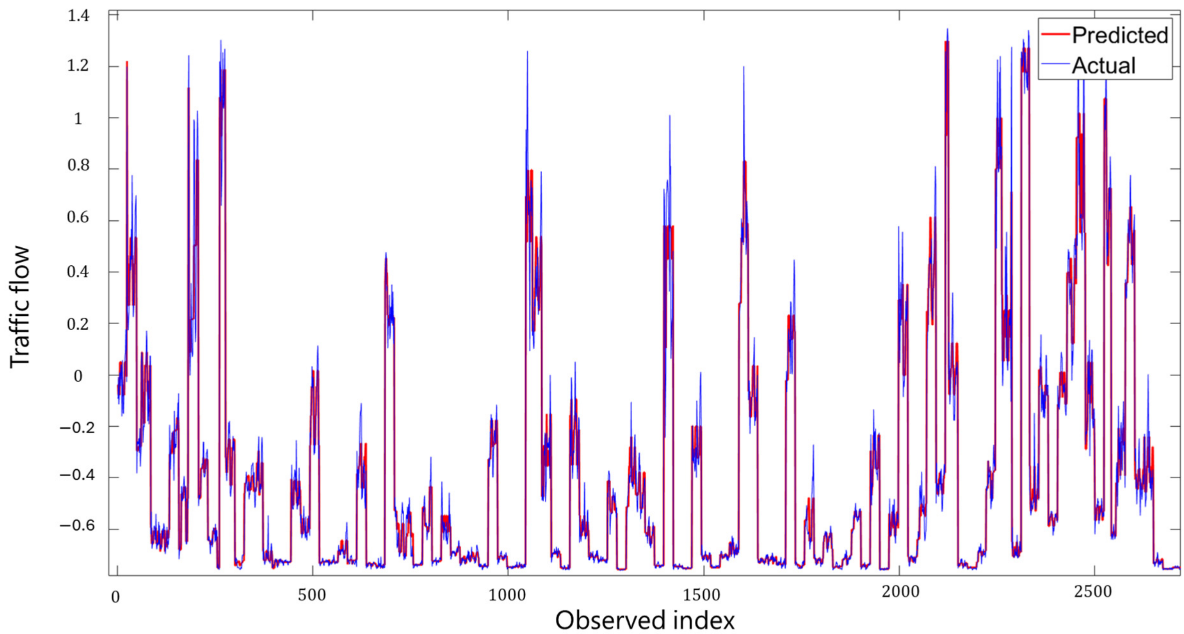


**Figure 4.** Plotting the prediction by Fine tree traffic-flow prediction.

### *4.2. Machine Learning: Classification and Density Estimation*

ML approaches were also used for predicting the traffic density. The performance of these classification methods was also examined using Scenarios 1–5 (Table 3). The results suggest that the addition of urban features also help improve the performance of the classification phase, in that Scenario 5 returned the highest level of accuracy with a 98.7% accuracy with Fine KNN (Table 5). Between the five scenarios, Scenario 1 yielded consistently low performances across all ML methods applied. For instance, the Bagged tree model within the Ensemble family achieved a singularly high cross-validation rate of 78.73% in Scenario 1, and it showed even higher estimation performance for Scenarios 2–5. In other words, using the road characteristics and spatial features could lead to considerable improvements.

**Table 5.** Performance of traffic density classification by ML models under 5 scenarios.

| ML-Model \ Scenario | Traffic Indic. (TI) | Road Chara. (RC) | Public Facility (POI) | RC and POI | All Features (TI, RC, POI) |
|---|---|---|---|---|---|
| Decision Tree (Fine) | 69.89 | 84.34 | 87.35 | 81.78 | 89.14 |
| Decision Tree (Med.) | 65.15 | 69.09 | 69.21 | 68.74 | 71.46 |
| Kernel Naïve Bayes | 56.27 | 57.74 | 64.50 | 57.72 | 64.37 |
| SVM (Quadratic) | 37.53 | 59.27 | 63.10 | 58.62 | 70.24 |
| SVM (Cubic) | 32.95 | 58.87 | 57.59 | 61.49 | 92.82 |
| SVM (Fine Gaussian) | 65.93 | 81.44 | 86.40 | 78.91 | 96.81 |
| SVM (Med. Gaussian) | 62.63 | 66.68 | 67.57 | 70.29 | 76.29 |
| KNN (Fine) | 65.81 | 88.77 | 87.12 | 91.72 | 98.70 |
| KNN (Medium) | 60.83 | 77.48 | 77.75 | 80.49 | 88.12 |
| KNN (Cos) | 59.78 | 76.41 | 76.32 | 64.21 | 88.07 |

**Table 5.** *Cont.*

| Scenario / ML-Model | Traffic Indic. (TI) | Road Chara. (RC) | Public Facility (POI) | RC and POI | All Features (TI, RC, POI) |
|---|---|---|---|---|---|
| KNN (Cubic) | 60.34 | 75.28 | 75.33 | 80.61 | 85.04 |
| KNN (Weighted) | 66.50 | 88.95 | 86.06 | 77.90 | 97.16 |
| Ensemble (Bagged) | 78.73 | 94.61 | 91.67 | 82.86 | 98.67 |
| KNN (Subspace) | 25.17 | 82.77 | 92.46 | 71.76 | 98.62 |
| Neural Network (Narrow) | 58.05 | 68.56 | 67.96 | 67.24 | 71.04 |
| Neural Network (Medium) | 59.62 | 79.48 | 77.56 | 89.34 | 92.75 |
| Neural Network (Wide) | 64.82 | 95.42 | 94.38 | 85.63 | 98.24 |
| Neural Network (bilayered) | 60.07 | 75.86 | 76.03 | 80.86 | 83.88 |
| Neural Network (trilayered) | 62.08 | 79.13 | 82.14 | 85.75 | 90.44 |

SVM: Support Vector Machine; KNN: K-Nearest Neighbour.

In all five scenarios, Fine KNN, Bagged tree, and Wide NN consistently returned a strong performance. Fine KNN achieved the highest accuracy of all MLs applied in Scenarios 4 and 5, which again confirmed the positive impact the road characteristics and the POI features had on the accuracy of the model.

Figure 5 shows the classification confusion matrix for the Fine KNN estimation model in Scenario 5. The model yielded a consistently strong performance, accurately classifying 98.3% of the 10,692 cases into their respective traffic density class (1032 cases into class 1: low density; 6059 into class 2: medium density; and 3420 cases into class 3: high density), whilst containing the volume of all incorrect outputs at 1.7% (181 out of 10,692 cases).

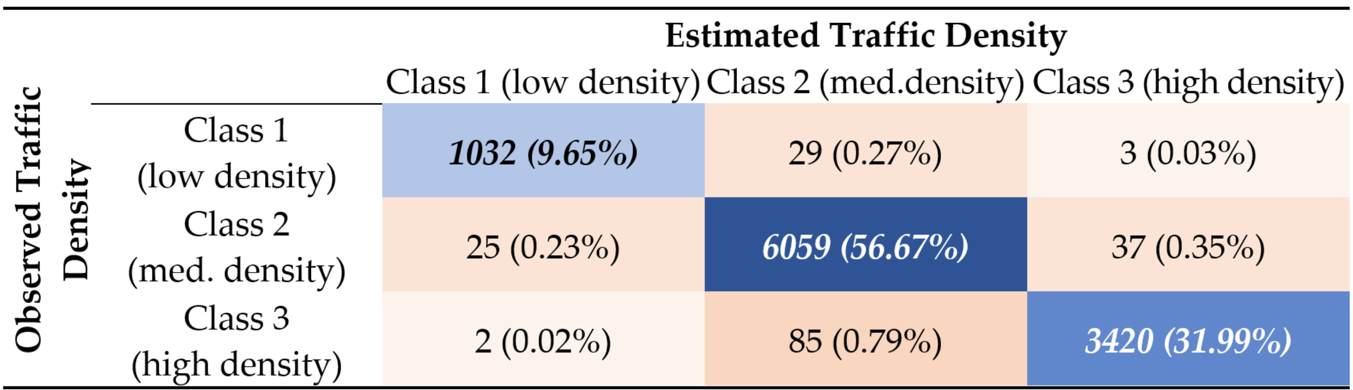


**Figure 5.** Classification confusion matrix derived with the Fine KNN model.

## 5. Sensitivity Analysis

### *5.1. Investigating the Sensitivity of the Kernel Density Bandwidth*

As mentioned earlier, use of urban features in ML-based traffic prediction is still growing and for those that employ KDE for POIs or similar, the sensitivity of their bandwidth is yet to be explored in a systematic fashion. The outcomes of the ML-based regression and classification tests in our study demonstrate that the KDE of the proposed urban features made a considerable improvement in their prediction accuracy. Yet, the KDE bandwidth itself was fixed at 1000 m, and we did not explore how the models would perform against different KDE bandwidth.

The bandwidth of KDE can be considered as the sphere of influence of each feature, and its search radius would naturally affect the outcome. A shorter bandwidth will retain more localised tendencies but may result in a noisy and abrupt distribution which may not be an accurate portrayal of the traffic flow; whilst a longer bandwidth will generate a smoother surface but with less information on the local traffic situation, as it is averaged out. To find the right balance between sensitivity to the local situation and the overall smoothness, sensitivity analysis of KDE becomes crucial. This study pursues it in three steps: (1) investigate the extent of changes in the model performance when using different bandwidths in KDE, (2) determine the optimal bandwidth for achieving the highest accuracy of traffic prediction, and (3) measure the impact of bandwidth on the accuracy of the ML models.

### 5.2. Sensitivity Analysis of the KDE Bandwidth

To test the sensitivity and the performance of the ML models against different KDE bandwidths, we will use four of the ML models that yielded the highest accuracy at 1000 m and conduct sensitivity analysis under Scenario 5 (i.e., all urban features and road characteristics) with bandwidth values of 200, 600, 1000, 1500, and 2000 m.

As the optimal KDE bandwidth is a function of the POI density and distribution, it may vary by the configuration of the study area as well as the dataset used. However, the following exercise should give us sufficient insights into how sensitive the ML models are against changes in the KDE bandwidth.

Table 6 summarises the results of the top-performing ML regression models (Fine tree, Wide Neural Network, Bagged tree, and Medium tree) with the five bandwidth distances. The results generally show that the choice of the KDE bandwidth has some influence on the model's validation metrics (*RMSE*, *MAE*, and $R^2$). However, the extent of their influence remains marginal in some models. The results also confirm that the use of 1000 m KDE bandwidth extracts the highest performance from these ML models for this dataset.

**Table 6.** Sensitivity analysis of the KDE bandwidth for ML traffic-flow prediction.

| Model | Bandwidth | *RMSE* | *R-Squared* | *MAE* |
|---|---|---|---|---|
| Fine tree | 200 | 0.160 | 0.88 | 0.082 |
| | 600 | 0.132 | 0.92 | 0.071 |
| | 1000 | 0.129 | 0.94 | 0.069 |
| | 1500 | 0.133 | 0.92 | 0.071 |
| | 2000 | 0.131 | 0.92 | 0.071 |
| Wide Neural Network | 200 | 0.187 | 0.84 | 0.098 |
| | 600 | 0.169 | 0.87 | 0.112 |
| | 1000 | 0.129 | 0.92 | 0.089 |
| | 1500 | 0.134 | 0.90 | 0.092 |
| | 2000 | 0.165 | 0.88 | 0.110 |
| Bagged tree | 200 | 0.146 | 0.90 | 0.083 |
| | 600 | 0.145 | 0.90 | 0.075 |
| | 1000 | 0.130 | 0.92 | 0.074 |
| | 1500 | 0.133 | 0.91 | 0.075 |
| | 2000 | 0.132 | 0.91 | 0.075 |
| Medium tree | 200 | 0.164 | 0.88 | 0.087 |
| | 600 | 0.147 | 0.90 | 0.083 |
| | 1000 | 0.145 | 0.91 | 0.082 |
| | 1500 | 0.146 | 0.90 | 0.084 |
| | 2000 | 0.148 | 0.90 | 0.084 |

Figure 6 shows the validation metrics for the best four ML models across different values of bandwidth. Each model exhibited a generally similar tendency of changes against the KDE bandwidth values. Among them, the Medium and Bagged trees showed low variation in the validation metrics. The Bagged tree, in particular, showed very little variation in its validation metrics across different KDE bandwidths, where $R^2$ converged between 0.90 and 0.92, *RMSE* between 0.130 and 0.146, and *MAE* between 0.074 and 0.083. Medium tree validation metrics also showed a stable tendency across different bandwidths, with $R^2$ ranging between 0.88 and 0.91, *RMSE* between 0.145 and 0.164, and *MAE* between 0.082 and 0.087.

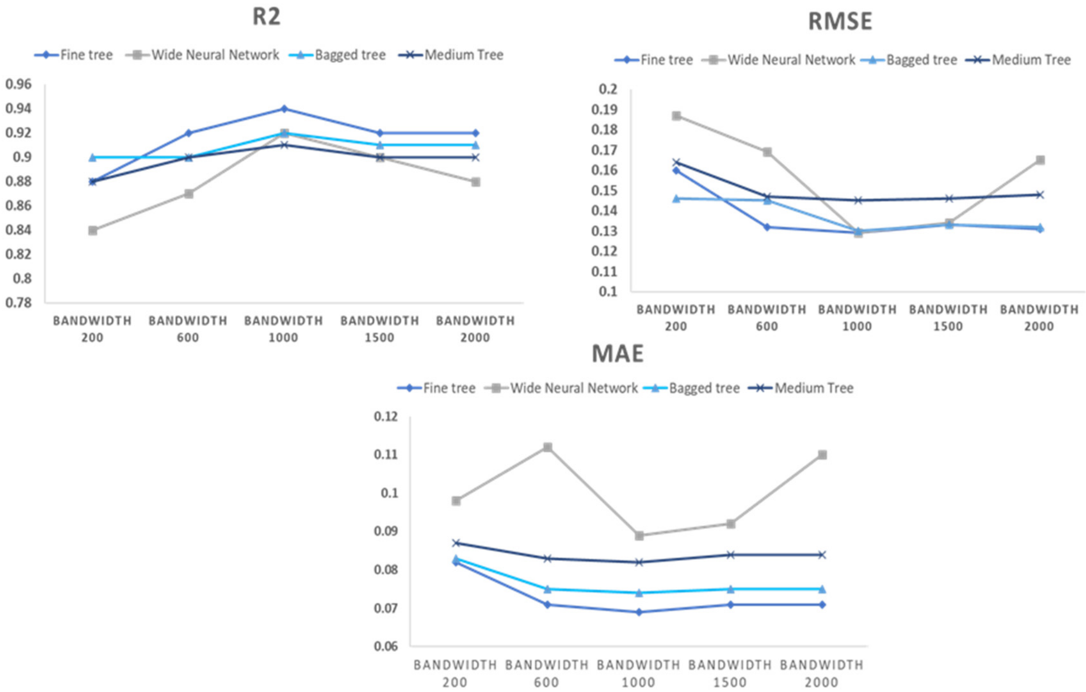


**Figure 6.** ML regression traffic-flow prediction models' validation metrics.

In contrast, Fine tree and Wide NN showed more volatility against changes in the bandwidth. Wide NN has the highest fluctuation in the validation metrics between the models, with the $R^2$ value ranging between 0.84 and 0.92.

### *5.3. Sensitivity Analysis of the Classification Models*

Table 7 summarises the results of ML classification models with the highest performance returned by Fine KNN, Bagged tree, Subspace KNN, and Wide Neural Network. They were also applied across the same five values of bandwidths (200 m, 600 m, 1000 m, 1500 m, and 2000 m). The results are similar to those from the ML regression modelling, where the bandwidth of 1000 m returned the highest performance from all tested models, and changes in the KDE bandwidth made some, if not drastic, impact on their validation metrics (*RMSE*, *MAE*, and $R^2$). Figure 6 illustrates changes in the cross-validation performance in estimating traffic density by these ML models across the bandwidths tested. In most cases, the models show significant improvement in cross-validation accuracy as the bandwidth increases from 200 m to 1000 m and a moderate decline after 1000 m. Subspace KNN and Bagged tree are more sensitive than other ML methods are against changes in the bandwidth, and their level of accuracy increased from 73.9% to 98.6%, and from 75.3% to 98.67%, respectively.

**Table 7.** Sensitivity analysis for ML traffic density prediction models.

| Model | Bandwidth | Cross-Validation Accuracy (%) |
|---|---|---|
| Fine KNN | 200 | 85.24 |
| | 600 | 90.69 |
| | 1000 | 98.70 |
| | 1500 | 93.73 |
| | 2000 | 91.28 |
| Bagged tree | 200 | 75.32 |
| | 600 | 91.40 |
| | 1000 | 98.67 |
| | 1500 | 96.10 |
| | 2000 | 94.37 |
| Subspace KNN | 200 | 73.95 |
| | 600 | 79.74 |
| | 1000 | 98.62 |
| | 1500 | 94.26 |
| | 2000 | 94.77 |
| Wide Neural Network | 200 | 83.64 |
| | 600 | 88.83 |
| | 1000 | 98.24 |
| | 1500 | 95.30 |
| | 2000 | 96.15 |

The results also suggest that the ranking of the performance of ML models changes between different bandwidth values (Figure 7). For example, Bagged tree provides higher accuracy than Wide KNN and Fine KNN at 600 m bandwidth, whereas, at 200 m bandwidth, Bagged tree provides lower accuracy than Wide KNN and Fine KNN. The choice of the optimal model would depend on the bandwidth values, which in turn may be confined by the specific dataset and the study area used.

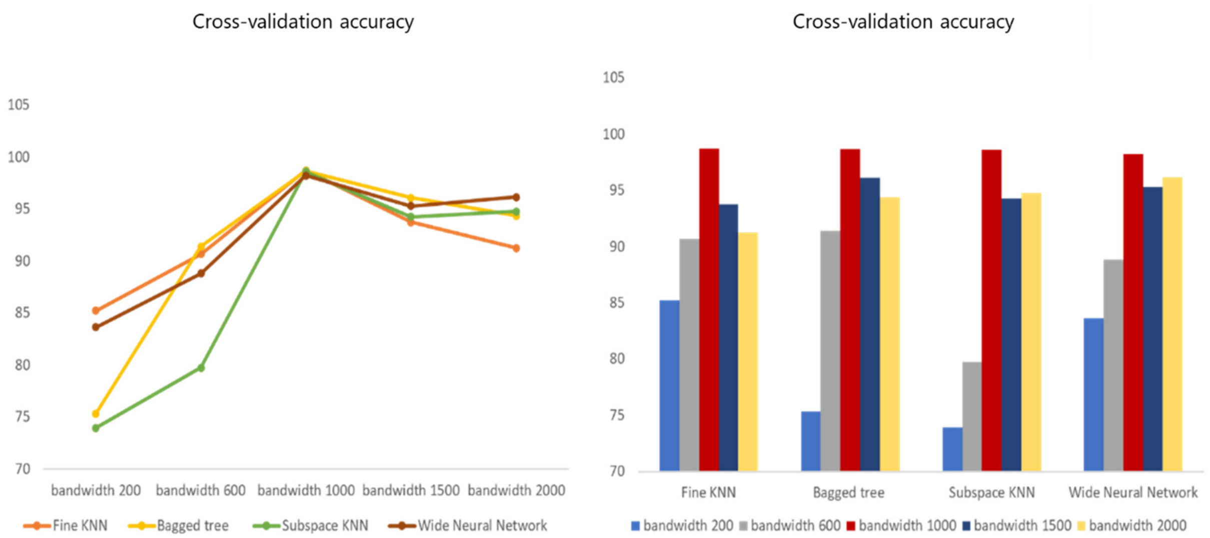


**Figure 7.** Changes in the cross-validation accuracy across five different KDE bandwidths for the highest-performing ML classification models.

Results from the sensitivity analysis of ML regression and ML classification suggest that (1) 1000 m was the best overall bandwidth among those used for sensitivity analysis (200, 600, 1000, 1500, and 2000 m), both for the regression and the classification models, and (2) classification is more sensitive to changes in the KDE bandwidth, thus highlighting the importance of identifying a suitable bandwidth for the classification models.

## 6. Discussion

The main goal of this study was to assess the importance of urban and spatial features, how their presence would affect the local traffic flow, and whether we can improve the traffic-flow prediction and traffic density estimation by accounting for the proximity to these facilities [18,19]. As illustrated in the previous section, the urban features had the highest impact among all other predictors towards improving the performance of the model. For instance, the highest-performing ML regression model saw its $R^2$ value rise from 0.59 to 0.91, whilst the *RMSE* and the *MAE* values reduced from 0.303 to 0.145, and from 0.211 to 0.018, respectively. In addition, the accuracy of the ML classification method itself has yielded a much higher performance, improving its forecasting from 65.81% to 98.7%, by utilising urban features besides the basic traffic indicators. These results demonstrate the importance of spatial features that could impact the traffic flow to improve the prediction models. In other words, we should reflect on the extent each region or urban area is affected by their specific urban factors that may affect the local traffic flow.

In our study, ML methods demonstrated their capacity to produce a robust and accurate prediction whilst also handling complex non-linear relationships between features and responses, thus uncovering the intricate patterns of dependencies that may have been otherwise obscured. Results from the prediction analysis of ML methods also demonstrated that their performance was greatly enhanced when urban features were introduced as part of the predictors. The best ML regression model was achieved by Fine tree using all proposed features with validation metrics ($R^2$ = 0.94, *RMSE* = 0.129, and *MAE* = 0.069), while the validation metrics of OLS regression and spatial lags and spatial errors were consistently lower at ($R^2$ = 0.30, *RMSE* = 0.397, and *MAE* = 0.288), ( $R^2$ = 0.43, *RMSE* = 0.388, and *MAE* = 0.282), and ($R^2$ =0.44, *RMSE* = 0.358, and *MAE* = 0.280), respectively. It confirms the comparative advantage of ML models for making traffic prediction, which aligns well with the existing literature [24,25]. Interestingly, the use of urban features would further enhance the performance of ML models but has little impact on the performance of non-ML regression models.

The fact that the Decision tree family, especially Fine tree, returned the best overall performance is rather unsurprising, given its affinity with multiple data (i.e., traffic characteristics and urban features) as well as its flexibility that facilitates classification and regression through data mining. It is also less dependent on large volume or high quality of data, thus making it a suitable and robust choice for traffic prediction that incorporates proximity to urban facilities.

Since the proximity to urban features and public space plays a key role in our ML traffic prediction, the choice of KDE bandwidth requires careful consideration, as does the rationale for using KDE itself. In principle, KDE allows us to smooth out the sphere of influence of POIs in all directions, and it offers a sensible solution for reflecting the impact of urban facilities on local traffic, especially when we have no data on the weighting or the significance of each facility, or the directions of flows around them. Where appropriate, KDE also allows us to incorporate weighting of each POI (e.g., estimated annual visitor counts). On the downside, KDE renders a simple smoothing technique, and its representation is neither statistically verified, nor does it specify the exact value as the most suitable bandwidth for traffic modelling. Still, it offers a suitable representation of point density in the area, and it is for this reason that the prevailing literature utilises KDE widely [11,15]. Our study aligns with this trend whilst also contributing to the literature by adding a systematic account of the sensitivity of the KDE bandwidth.

Results from sensitivity analysis show that the performance of the ML regression models and that of the classification models is affected by the choice of KDE bandwidth, with classification being more significantly affected by the change in the KDE bandwidth. Overall, 1000 m was the optimal bandwidth for both regressions and classification models. It is unclear as to why 1000 m was the most effective representation of such urban features. We would hypothesise that 1000 m marks the distance for the sphere of influence, or more specifically, the area within which local travel would increase from a transport junction (e.g., train station, bus stops) to the destination, whilst for private vehicles, more parking and loading/unloading activities might arise, thus generating local traffic congestion. This needs further ground truthing and empirical investigation.

Regardless, the optimal value is a function of POI density and geometry, which means that the value that constitutes the optimal bandwidth may change from one study area and dataset to another. The only viable solution would be to carry out sensitivity analysis first to define the optimal KDE bandwidth for the particular study area empirically which, in the case of our study, was 1000 m radius. This confirms the indicative examples identified in the literature, thus suggesting its suitability across several different contexts and datasets.

The proposed framework has several limitations. First, sensitivity of the KDE was tested with a finite set of discrete bandwidths. It was also applied to all POIs and public transport facilities to maintain simplicity and consistency. Since these spatial features may trigger various traffic behaviours, future works should reassess the bandwidth for each feature, as this may further improve the performance of the models.

The size of the dataset also needs attention. The traffic data used in this study consisted of 10,692 rows, but a larger dataset would help refine the models further. The ways in which the dataset is prepared and processed also requires careful consideration. As our study focused on developing and evaluating the performance of an offline model that is designed for specific applications (which have their own utility, e.g., urban design and transport planning), the annual estimate of traffic flows adopted was perfectly suitable. However, as a future avenue of our research, we invariably wish to make near-real-time traffic prediction. In this context, it would be vitally important to use a larger dataset whilst also accounting for any missing entries or anomalies and sudden changes in traffic conditions (e.g., those triggered by traffic accidents nearby). The framework and the predictors proposed in this study should help inform such future works on developing DL real-time traffic flow prediction using the features proposed in this study, but their sensitivity to such real-world data requires further sensitivity analysis.

## 7. Conclusions

This study aimed to enhance and evaluate the performance of ML-based traffic prediction models by comparing between 21 ML and non-ML models, and applying them to different combinations of predictors, including proximity to urban features. Sensitivity analysis was also carried out to determine the most suitable bandwidth for KDE. The results show that the performance of ML methods is greatly enhanced by adding more predictors, especially the spatial features. The change in KDE bandwidth also highlighted the importance of deriving a suitable value to maximise the performance of the ML model.

The study also developed three non-ML traffic prediction models (OLS, spatial error, and spatial lag), which were applied to the same set of features and dataset as those utilised for the ML models. The results confirmed that ML methods were comprehensively more accurate than these statistical methods in predicting the traffic flow. The best regression ML prediction was achieved by Fine tree with validation metrics of $R^2$ = 0.94, $RMSE$ = 0.129, and $MAE$ = 0.069; whereas the classification ML prediction for traffic density projection was championed by Fine KNN with a cross-validation accuracy of 98.7%.

Our contribution was to demonstrate that the combination of a select few ML models with proximity to urban features can offer a remarkably high level of precision in predicting the traffic flow as well as estimating the traffic density. While our model was designed as an offline model, there is a distinct scope for developing a model for accurate, real-time prediction. However, as the use of urban features is crucial for both regression ML and classification ML, the specific challenge lies with the preparation of a suitable set of data to support the real-time projection of the traffic situation. In other words, a real-time prediction model will require resilience against unprocessed real-world data with missing entries and abrupt changes to the traffic situation, and further investigation is wanted for developing models that can persevere in those conditions.

**Author Contributions:** Conceptualisation, R.R. and S.S.; methodology, R.R. and S.S.; validation, R.R., S.S. and N.S.; formal analysis, R.R. and S.S.; investigation, R.R.; data curation, R.R.; writing—original draft preparation, R.R.; writing—review and editing, N.S. and S.S.; visualisation, R.R.; supervision, S.S.; project administration, R.R., S.S. and N.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data on road characteristics and traffic flow used in this study are publicly available from Roadtraffic.DFT.gov.uk (UK Department for Transport 2022). Road Traffic Statistics—London Region. Available online: https://roadtraffic.dft.gov.uk/regions/6 (accessed on 24 April 2023). The Point of Interest data (including hospitals, colleges, recreational places, tourism locations, and landscape areas) are publicly available from Edina Digimap. Available online: https://digimap.edina.ac.uk/ (accessed on 24 April 2023).

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Paul, A.; Chilamkurti, N.; Daniel, A.; Rho, S. *Intelligent Vehicular Networks and Communications: Fundamentals, Architectures and Solutions*; Elsevier: Amsterdam, The Netherlands, 2016.
2. Miglani, A.; Kumar, N. Deep learning models for traffic flow prediction in autonomous vehicles: A review, solutions, and challenges. *Veh. Commun.* **2019**, *20*, 100184. [CrossRef]
3. Melakessou, F.; Derrmann, T.; Engel, T. Asymmetry analysis of inbound/outbound car traffic load distribution in Luxembourg. In Proceedings of the 13th ACM International Symposium on Mobility Management and Wireless Access, Cancun, Mexico, 2–6 November 2015; pp. 5–12.
4. World Health Organization. Global Status Report on Road Safety 2018 Summary. Available online: https://www.who.int/publications/i/item/9789241565684 (accessed on 30 June 2024).
5. Shaygan, M.; Meese, C.; Li, W.; Zhao, X.; Nejad, M. Traffic prediction using artificial intelligence: Review of recent advances and emerging opportunities. *Transp. Res. Part C Emerg. Technol.* **2022**, *145*, 103921. [CrossRef]
6. Wu, Y.; Tan, H.; Qin, L.; Ran, B.; Jiang, Z. A hybrid deep learning based traffic flow prediction method and its understanding. *Transp. Res. Part C Emerg. Technol.* **2018**, *90*, 166–180. [CrossRef]
7. Xie, P.; Li, T.; Liu, J.; Du, S.; Yang, X.; Zhang, J. Urban flow prediction from spatiotemporal data using machine learning: A survey. *Inf. Fusion* **2020**, *59*, 1–12. [CrossRef]
8. Tafidis, P.; Teixeira, J.; Bahmankhah, B.; Macedo, E.; Coelho, M.C.; Bandeira, J. Exploring crowdsourcing information to predict trafficrelated impacts. In Proceedings of the 2017 IEEE International Conference on Environment and Electrical Engineering and 2017 IEEE Industrial and Commercial Power Systems Europe (EEEIC/I&CPS Europe), Milan, Italy, 6–9 June 2017; pp. 1–6.
9. Han, L.; Huang, Y.S. Short-term traffic flow prediction of road network based on deep learning. *IET Intell. Transp. Syst.* **2020**, *14*, 495–503. [CrossRef]

10. Xu, Y.; Kong, Q.J.; Lin, S.; Liu, Y. Urban traffic flow prediction based on road network model. In Proceedings of the 2012 9th IEEE International Conference on Networking, Sensing and Control, Beijing, China, 11–14 April 2012; pp. 334–339.
11. Zeng, H.; Peng, Z.; Huang, X.; Yang, Y.; Hu, R. Deep spatio-temporal neural network based on interactive attention for traffic flow prediction. *Appl. Intell.* **2022**, *52*, 10285–10296. [CrossRef]
12. Long, K.; Lin, Q.; Gu, J.; Wu, W.; Han, L.D. Exploring traffic congestion on urban expressways considering drivers' unreasonable behavior at merge/diverge sections in China. *Sustainability* **2018**, *10*, 4359. [CrossRef]
13. Hou, Y.; Chen, J.; Wen, S. The effect of the dataset on evaluating urban traffic prediction. *Alex. Eng. J.* **2021**, *60*, 597–613. [CrossRef]
14. Shen, T.; Hong, Y.; Thompson, M.M.; Liu, J.; Huo, X.; Wu, L. How does parking availability interplay with the land use and affect traffic congestion in urban areas? The case study of Xi'an, China. *Sustain. Cities Soc.* **2020**, *57*, 102126. [CrossRef]
15. Jia, R.; Khadka, A.; Kim, I. Traffic crash analysis with point-of-interest spatial clustering. *Accid. Anal. Prev.* **2018**, *121*, 223–230. [CrossRef]
16. Fang, S.; Prinet, V.; Chang, J.; Werman, M.; Zhang, C.; Xiang, S.; Pan, C. MS-Net: Multi-source spatio-temporal network for traffic flow prediction. *IEEE Trans. Intell. Transp. Syst.* **2021**, *23*, 7142–7155. [CrossRef]
17. Xiao, W.; Kuang, L.; An, Y. Traffic flow prediction through the fusion of spatial-temporal data and points of interest. In Proceedings of the Database and Expert Systems Applications: 32nd International Conference, DEXA 2021, Virtual Event, 27–30 September 2021; Part I. pp. 314–327.
18. James, J. Citywide traffic speed prediction: A geometric deep learning approach. *Knowl.-Based Syst.* **2021**, *212*, 106592.
19. Li, Y.; Yu, R.; Shahabi, C.; Liu, Y. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. *arXiv* **2017**, arXiv:1707.01926.
20. George, S.; Santra, A.K. Traffic Prediction Using Multifaceted Techniques: A Survey. *Wirel. Pers. Commun.* **2020**, *115*, 1047–1106. [CrossRef]
21. Stathopoulos, A.; Karlaftis, M.G. A multivariate state space approach for urban traffic flow modeling and prediction. *Transp. Res. Part C Emerg. Technol.* **2003**, *11*, 121–135. [CrossRef]
22. Xia, Z.; Wu, J.; Wu, L.; Chen, Y.; Yang, J.; Yu, P.S. A Comprehensive Survey of the Key Technologies and Challenges Surrounding Vehicular Ad Hoc Networks. *ACM Trans. Intell. Syst. Technol.* **2021**, *12*, 37. [CrossRef]
23. Li, R.; Pereira, F.C.; Ben-Akiva, M.E. Overview of traffic incident duration analysis and prediction. *Eur. Transp. Res. Rev.* **2018**, *10*, 22. [CrossRef]
24. Zantalis, F.; Koulouras, G.; Karabetsos, S.; Kandris, D. A review of machine learning and IoT in smart transportation. *Future Internet* **2019**, *11*, 94. [CrossRef]
25. Cui, Z.; Henrickson, K.; Ke, R.; Wang, Y. Traffic graph convolutional recurrent neural network: A deep learning framework for network-scale traffic learning and forecasting. *IEEE Trans. Intell. Transp. Syst.* **2020**, *21*, 4883–4894. [CrossRef]
26. Khan, N.U.; Shah, M.A.; Maple, C.; Ahmed, E.; Asghar, N. Traffic flow prediction: An intelligent scheme for forecasting traffic flow using air pollution data in smart cities with bagging ensemble. *Sustainability* **2022**, *14*, 4164. [CrossRef]
27. Harrou, F.; Zeroual, A.; Sun, Y. Traffic congestion monitoring using an improved kNN strategy. *Measurement* **2020**, *156*, 107534. [CrossRef]
28. Zhang, L.; Liu, Q.; Yang, W.; Wei, N.; Dong, D. An Improved K-nearest Neighbor Model for Short-term Traffic Flow Prediction. *Procedia-Soc. Behav. Sci.* **2013**, *96*, 653–662. [CrossRef]
29. Smith, B.L.; Williams, B.M.; Oswald, R.K. Comparison of parametric and nonparametric models for traffic flow forecasting. *Transp. Res. Part C Emerg. Technol.* **2002**, *10*, 303–321. [CrossRef]
30. Chandra, S.R.; Al-Deek, H. Predictions of freeway traffic speeds and volumes using vector autoregressive models. *J. Intell. Transp. Syst.* **2009**, *13*, 53–72. [CrossRef]
31. Lu, Z.; Zhou, C.; Wu, J.; Jiang, H.; Cui, S. Integrating granger causality and vector auto-regression for traffic prediction of large-scale WLANs. *KSII Trans. Internet Inf. Syst.* **2016**, *10*, 136–151.
32. Williams, B.M.; Hoel, L.A. Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results. *J. Transp. Eng.* **2003**, *129*, 664–672. [CrossRef]
33. Shekhar, S.; Williams, B.M. Adaptive seasonal time series models for forecasting short-term traffic flow. *Transp. Res. Rec. J. Transp. Res. Board* **2007**, *2024*, 116–125. [CrossRef]
34. Kashyap, A.A.; Raviraj, S.; Devarakonda, A.; Nayak-K, S.R.; Santhosh, K.V.; Bhat, S.J. Traffic flow prediction models—A review of deep learning techniques. *Cogent Eng.* **2022**, *9*, 2010510. [CrossRef]
35. Medina-Salgado, B.; Sánchez-Delacruz, E.; Pozos-Parra, P.; Sierra, J.E. Urban traffic flow prediction techniques: A review. *Sustain. Comput. Inform. Syst.* **2022**, *35*, 100739. [CrossRef]
36. Yang, X.; Zou, Y.; Tang, J.; Liang, J.; Ijaz, M. Evaluation of short-term freeway speed prediction based on periodic analysis using statistical models and machine learning models. *J. Adv. Transp.* **2020**, *2020*, 16. [CrossRef]
37. Ramchandra, N.R.; Rajabhushanam, C. Machine learning algorithms performance evaluation in traffic flow prediction. *Mater. Today Proc.* **2022**, *51*, 1046–1050. [CrossRef]

38. Ara, Z.; Hashemi, M. Identifying the severity of road accident impact on traffic flow by ensemble model. In Proceedings of the IEEE the 22nd International Conference on Information Reuse and Integration for Data Science (IRI), Las Vegas, NV, USA, 10–12 August 2021; pp. 115–122.
39. Sun, P.; Aljeri, N.; Boukerche, A. Machine learning-based models for realtime traffic flow prediction in vehicular networks. *IEEE Netw.* **2020**, *34*, 178–185. [CrossRef]
40. Wang, J.; Chen, R.; He, Z. Traffic speed prediction for urban transportation network: A path based deep learning approach. *Transp. Res. Part C Emerg. Technol.* **2019**, *100*, 372–385. [CrossRef]
41. Alajali, W.; Zhou, W.; Wen, S.; Wang, Y. Intersection traffic prediction using decision tree models. *Symmetry* **2018**, *10*, 386. [CrossRef]
42. Liu, Y.; Zhang, N.; Luo, X.; Yang, M. Traffic Flow Forecasting Analysis based on Two Methods. *J. Phys. Conf. Ser.* **2021**, *1861*, 012042. [CrossRef]
43. Liu, Z.; Lyu, C.; Huo, J.; Wang, S.; Chen, J. Gaussian process regression for transportation system estimation and prediction problems: The Deformation and a Hat Kernel. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 2233122342. [CrossRef]
44. Wang, W.; Zhou, C.; He, H.; Wu, W.; Zhuang, W.; Shen, X. Cellular traffic load prediction with LSTM and Gaussian process regression. In Proceedings of the ICC 2020—2020 IEEE International Conference on Communications (ICC), Dublin, Ireland, 7–11 June 2020; pp. 1–6.
45. Ali, U.; Mahmood, T. Using Deep Learning to predict short term traffic flow: A systematic literature review. *Lect. Notes Inst. Comput. Sci. Soc. Inform. Telecommun. Eng.* **2018**, *222*, 90–101.
46. Wang, P.; Li, L.; Jin, Y.; Wang, G. Detection of unwanted traffic congestion based on existing surveillance system using in freeway via a CNN-architecture traffic net. In Proceedings of the 13th IEEE Conference on Industrial Electronics and Applications (ICIEA) 2018, Wuhan, China, 31 May–2 June 2018; pp. 1134–1139.
47. Yin, X.; Wu, G.; Wei, J.; Shen, Y.; Qi, H.; Yin, B. Deep learning on traffic prediction: Methods, analysis, and future directions. *IEEE Trans. Intell. Transp. Syst.* **2021**, *23*, 4927–4943. [CrossRef]
48. Zheng, H.; Lin, F.; Feng, X.; Chen, Y. A hybrid deep learning model with attention-based conv-LSTM networks for short-term traffic flow prediction. *IEEE Trans. Intell. Transp. Syst.* **2020**, *22*, 6910–6920. [CrossRef]
49. Defferrard, M.; Bresson, X.; Vandergheynst, P. Convolutional neural networks on graphs with fast localized spectral filtering. *arXiv* **2016**, arXiv:1606.09375.
50. Fang, S.-H.; Fei, Y.-X.; Xu, Z.; Tsao, Y. Learning transportation modes from smartphone sensors based on deep neural network. *IEEE Sens. J.* **2017**, *17*, 61116118. [CrossRef]
51. Qu, L.; Li, W.; Li, W.; Ma, D.; Wang, Y. Daily long-term traffic flow forecasting based on a deep neural network. *Expert Syst. Appl.* **2019**, *121*, 304–312. [CrossRef]
52. Parsa, A.B.; Shabanpour, R.; Mohammadian, A.; Auld, J.; Stephens, T. A data-driven approach to characterize the impact of connected and autonomous vehicles on traffic flow. *Transp. Lett.* **2021**, *13*, 687–695. [CrossRef]
53. Fang, S.; Zhang, Q.; Meng, G.; Xiang, S.; Pan, C. GSTNet: Global Spatial Temporal Network for traffic flow prediction. In Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence (IJCAI-19), Macao, China, 10–16 August 2019; pp. 2286–2293.
54. Nguyen-Phuoc, D.Q.; Young, W.; Currie, G.; De Gruyter, C. Traffic congestion relief associated with public transport: State-of-the-art. *Public Transp.* **2020**, *12*, 455–481. [CrossRef]
55. Zhao, Y.; Lin, Y.; Wen, H.; Wei, T.; Jin, X.; Wan, H. Spatial-temporal position-aware graph convolution networks for traffic flow forecasting. *IEEE Trans. Intell. Transp. Syst.* **2022**, *24*, 8650–8666. [CrossRef]
56. Rauch, J.; Hübner, U.; Denter, M.; Babitsch, B. Improving the prediction of emergency department crowding: A time series analysis including road traffic flow. In Proceedings of the 13th Health Informatics Meets Digital Health Conference, Vienna, Austria, 28–29 May 2019; pp. 57–64.