ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume X-4/W3-2022
7th International Conference on Smart Data and Smart Cities (SDSC), 19–21 October 2022, Sydney, Australia

PREDICTING TRAFFIC CONGESTION DURING COVID19 USING HUMAN MOBILITY AND STREET-WASTE FEATURES

N.Zarbakhsh [*], G.McArdle

School of Computer Science, University College Dublin, Dublin, Belfield, Dublin 4, Ireland
negin.zarbakhsh@ucdconnect.ie, gavin.mcardle@ucd.ie

Commission IV, WG IV/9

KEY WORDS: COVID-19, Feature Importance, Human Mobility, Machine Learning, Street-waste, Traffic Congestion.

ABSTRACT:

With COVID-19's prevalence and government efforts to curb its spread, urban travel behaviour has significantly altered, resulting in a significant shift in traffic congestion. Rather than predicting traffic congestion based on historical data, we aim to model the correlation between travel behaviour and external mobility-related urban features and use Dublin in Ireland as a case study. This study incorporates four categories of urban data, including 1) Mobility-based features, including the government's interventions and mobility pattern changes in different locations, 2) Environmental features such as weather and urban street-waste, 3) COVID-19-related features such as the positivity and vaccination rates, and 4) Time-related features such as public holidays. First, we examine the impact of COVID-19 on traffic congestion and street-waste to understand the city's dynamic. Then, multiple machine learning (ML) models, such as random forests, support vector regression, light gradient boosting machine, and multiple linear regression are trained, and their performance optimized to predict traffic congestion changes. We compare the outcomes of the models with several evaluation metrics and interpret the best performing model. The results indicate that mobility changes in grocery and pharmacy, retail and recreation, workplaces sectors, and the amount of urban street-waste significantly contribute to the model outcomes. Findings could predict traffic dynamics in times of crisis and allow authorities to comprehend the effects of their intervention measures on mobility, which would ultimately benefit developing smart cities and intelligent transportation systems.

1. INTRODUCTION

The COVID-19 coronavirus has spread globally with more than 410 million confirmed cases and more than 5.80 million deaths as of February 12, 2022, according to WHO data[1]. Human mobility is one of the main reasons for the rapid spread of COVID-19, which has been dramatically affected by government non-pharmaceutical interventions (NPIs) aimed at limiting social and economic activities. At the end of March 2020, around 50% of global road transport activities were below the 2019 average, and 75% of commercial flight activity was below the 2019 average by mid-April 2020 (Sung and Monschauer, 2020). During COVID-19, many people have started taking their cars, electric scooters, and bicycles instead of public transportation[2], which all resulted in a reduction in services provided by transportation systems. On top of this, many companies have resorted to remote working, and many places have shortened their opening hours, all of which have caused drastic changes in travel behaviour. The concept of traffic congestion, defined by a prolonged travel time compared to free-flowing traffic, is fundamental for urban planners to ensure a stable transportation system. Urban planners manage real-time traffic better by reducing congestion, controlling traffic lights, guiding routes to reduce traffic congestion, and rescheduling public transportation services. At the same time, taxi drivers and ride-sharing companies can plan their services more efficiently. Therefore, in this paper, we examine the problem of unexpected changes in traffic congestion patterns and propose an approach that incorporates human mobility characteristics into forecasting algorithms. Dublin, in Ireland, is used as a case study.

TomTom Traffic Index (TomTom, 2021), which measures historical and near-real-time traffic congestion levels, is one way to measure changes. This is an index that captures the extra amount of time that drivers are experiencing on the road. Dublin ranks 35th in the list of countries with the most traffic congestion. In Dublin, the congestion level in 2019 was 48%; in 2020 and 2021, it decreased to 38% and 36%, respectively. According to TomTom's definition, a 36% congestion level means travelers spent 36% longer travel time compared to an non-congested period. In January and February of 2019, traffic congestion in Dublin was around 46%; this figure increased to 50% in respective months in 2020; however, it declined to 38% in March 2020 following the COVID-19 pandemic. These figures continued to decline in 2020, with an average decrease of 30% and 27% in April and May, respectively [3]. According to these numbers, COVID-19 became such a significant threat in the second quarter of 2020 that the world and Dublin city began to shut down, and the traffic patterns shifted to unanticipated levels. Fig. 1 depicts the traffic congestion in Dublin on Fridays before the first confirmed case of COVID-19 in Ireland and after the government's first stay-at-home order on March 27, 2020. It can be seen that March experienced an unprecedented level of traffic congestion. The unpredictable changes in traffic congestion make traffic forecasting a challenging task, as most of the current algorithms rely heavily on the previous values and are incapable of predicting these changes. A Black Swan is a term used by Taleb (2007) to describe such events that occur unexpectedly, have significant consequences, and can only be fully understood after hindsight has been gained. Consequently, predictions that rely solely on previous values will have a sig-

* Corresponding author
[1] https://covid19.who.int/
[2] https://www.tomtom.com/en_gb/traffic-index/
[3] https://www.tomtom.com/en_gb/traffic-index/dublin-traffic#statistics

https://doi.org/10.5194/isprs-annals-X-4-W3-2022-301-2022 |

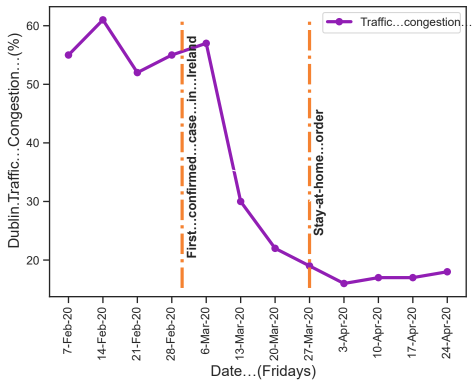


Figure 1. Traffic congestion in Dublin on Fridays before and after the first COVID-19 confirmed case in Ireland and the first stay-at-home order issued by the Irish government.

nificant forecasting error, and stationary statistical models are not appropriate in the presence of such anomalies. This encourages us to incorporate human-mobility-related features to capture human behaviour from different perspectives to predict future traffic congestion better. Understanding and predicting traffic congestion patterns is essential to managing and controlling non-recurrent traffic conditions during the COVID-19 pandemic, which can assist urban planners with future emergencies and pandemics.

Artificial Intelligence (AI) and the ability to store big spatiotemporal (ST) data are revolutionizing traffic studies. These developments are now being integrated into Intelligent Transportation Systems (ITS) to predict how people move and how this can affect traffic congestion. Predictions are often made using statistics and ML. Most statistical models are mathematically based, yet they struggle when faced with complex or nonlinear data (Karlaftis and Vlahogianni, 2011). By contrast, ML can handle vast amounts of data, has a flexible modelling capability, can generalize and adapt, and are generally good at prediction tasks (Karlaftis and Vlahogianni, 2011). In this study, we used four widely used ML models in traffic studies (Qiu and Fan, 2021; Zhang et al., 2018; Bratsas et al., 2020): Multiple Linear Regression (MLR), Random Forest (RF), Light Gradient Boosting (LGBM), and Support Vector Machine (SVR) to predict traffic congestion in Dublin city during COVID-19.

1.1 Contributions

Our approach incorporates novel features such as COVID-19-related features, mobility-based features, and urban street-waste features along with typical features such as weather and timing to improve traffic congestion prediction during this unprecedented situation. We also compared the results of all ML models using five evaluation metrics. We chose the best performing model based on the results of the five-fold cross-validation evaluation metrics. Finally, to produce an explainable ML model, we choose the best model and analyses the importance of the features based on SHapley Additive Explanations (SHAP). This work has the following main contributions:

- A detailed analysis of the impact of COVID-19 on Irish society by analyzing the relationship between human mobility and human travel behaviors.
- Exploring a key issue of urban computing: understanding traffic congestion based on ubiquitous urban data representing human movement and daily activities during the COVID-19 pandemic.
- Considering human mobility changes, NPIs, and street-waste for the first time as complementary features for traffic congestion prediction during the COVID-19.
- Developing an explainable ML model that explains the contribution of the novel features to the ML traffic congestion outcomes.

2. RELATED WORK

In the early stage of traffic studies, statistical methods, including time series models such as auto-regressive integrated moving average (ARIMA)- based models (Yu and Zhang, 2004), and Kalman filtering theory (Xie et al., 2007), Hidden Markov-based models (Qiao et al., 2018) were commonly used. Most of these methods assume linearity and stationary behaviour, which encourages researchers to use ML models to deal with traffic data that are nonlinear, dynamic, and contain spatiotemporal dependencies. Supervised ML approaches such as Support Vector Machine (SVM) (Zhang et al., 2018; Bratsas et al., 2020), Random Forest (Qiu and Fan, 2021; Bratsas et al., 2020), and Gradient Boost algorithms (Toan and Truong, 2021) are representative models used in urban traffic studies. In the following, we review state of the art and categorize the literature on traffic congestion before and after the COVID-19 outbreak.

Traffic congestion before COVID-19: Data on traffic sensors (Yang, 2013; Yu et al., 2017), taxis' GPS (Ma et al., 2015; Mridha et al., 2017) and social networking platforms are among the most commonly used variables for traffic studies. For traffic congestion prediction, Yang (2013) used the feature selection method was used to reduce the dimensionality of historical traffic sensor data and trained the probability-based scoring. Ma et al. (2015) gathered speed data collected from taxis' GPS and trained a deep Restricted Boltzmann Machine and Recurrent Neural Network architecture to reduce the high dimensionality. Yu et al. (2017) applied spatiotemporal recurrent convolutional networks based on sampling network-wide traffic speeds as a collection of static images, where each pixel represents a road segment's traffic condition. The proposed architecture captures both spatial and temporal dependencies by using Deep Convolutional Neural Networks (DCNNs) and Long-Short-Term Memory (LSTM) neural networks, which provide better accuracy for short-term and long-term predictions over other deep learning-based methods. As urban computing becomes more popular, studies of traffic with heterogeneous datasets become more popular. Song et al. (2016) collected big heterogeneous data, including transportation network and human mobility data, to develop a deep learning platform named DeepTransport, to predict human mobility and traffic congestion. Using ubiquitous data sets, Wu et al. (2016) analyzed traffic dynamics and predicted New York city taxi drop-offs using factors such as weather, regional functions, disasters, and vehicle collisions based on the Kernel Ridge Regression Degree-2 polynomial kernel. Based on tweet reports of road closures, Mridha et al. (2017) predicted taxi pickup hotspot location during various road closure incidents.

Traffic congestion after COVID-19: After the prevalence of COVID-19, anomalies occur in traffic due to changes in human travel behaviour as a result of government NPIs, which encourage researchers to collect more features to describe such sudden changes in mobility. Also, analysis of ST data has become one of the COVID-19-related topics to analyze the effects of NPI on human movement and individual travel behaviour (Kraemer et al., 2020; Li et al., 2021c). Before COVID-19, traffic anomalies were usually caused by accidents, traffic control, holidays, protests, sporting events, celebrations, natural disasters, (Wolshon and McArdle, 2009; Hara and Kuwahara, 2015) and other factors (Pan et al., 2013). For example, on a spatial and temporal scale, (Wolshon and McArdle, 2009) analyzed hurricane evacuation traffic patterns in southeast Louisiana. Hara and Kuwahara (2015) studied the evacuation behaviour in Japan and the traffic congestion following the earthquake using GPS data from smartphones and probe cars. While natural disasters tend to have a relatively short impact duration, the impact duration associated with the COVID-19 pandemic has been prolonged for more than two years, necessitating new models and features to explain such changes in travel behaviour. Using the traffic speed data collected from Baidu Maps, Li et al. (2021b) analyzed the empirical ST road congestion during the COVID-19 pandemic and employed the Singular Value Decomposition algorithm to capture the spatial and temporal variation. Li et al. (2021a), evaluate the impacts of NPIs on the use of public bicycle sharing in London using a segmented regression model with an interrupted time series approach. The results showed that the cycle hire in London decreased significantly after the first lockdown, whereas subsequent ease restrictions had no significant impacts. Additionally, there was a drastic change in the demand for bicycles near train stations and parks, which shows changes in human behaviour regarding bicycle usage. Also, ride-sharing services (e.g., Uber) witnessed tremendous changes in driver behaviour (Wang et al., 2022).

Literature shows that human behaviour and travel patterns have shifted significantly where traffic data is no longer stationary, and relying purely on historical data is not sufficient anymore. The long-term effects of the current pandemic necessitate the inclusion of novel features that quantify human activities into the traffic congestion forecasting pipeline to reduce the uncertainty in forecasting. Consequently, we introduce novel features such as mobility-based and urban street-waste features, which represent the level of congestion on streets, and typical traffic features to improve the forecasting accuracy during COVID-19 and similar crises that may occur in the future.

3. MATERIALS

In this section, our goal is to determine if mobility-based features can be used to predict tomorrow's traffic congestion by using an interpretable supervised ML model. We begin by reviewing the data, then the models, and finally the SHAP method.

3.1 Data Sources

COVID-19 Features: Data for COVID-19 positivity rate and daily number of vaccines are obtained from Ireland's COVID-19 Data Hub [4]. The Irish government's precaution measures are collected from Oxford COVID-19 Government Response Tracker (Hale et al., 2021). It contains restrictions including the closure of schools, and workplaces, cancellation of public events, limiting the public transport services, stay-at-home, restricting internal movements, and international travel controls.

Environment Features: Many studies have focused on the impacts of weather and, specifically, rain on traffic flow (Chen et al., 2019). To capture the impact of weather, we use the precipitation and minimum and maximum temperature data [5].

[4] https://covid19ireland-geohive.hub.arcgis.com/

[5] https://www.met.ie/climate/available-data/historical-data

Another environmental feature used in this study is a street cleaning data gathered by Thorntons on behalf of Dublin City Council [6] used as an indicator of human movements in cities. We also gathered biodegradable waste net weight (Ton) [7] to analyze the amount of waste in parks and compared it with mobility changes in parks.

Human Mobility Patterns: Daily country-level mobility changes were obtained from the COVID-19 Community Mobility site [8]. It provides data regarding the patterns of visit changes to places such as retail & recreation, grocery & pharmacy, parks, transit stations, workplaces, and residential areas compared to baseline days which is the median value for the five weeks, from January 3 to February 6, 2020.

Time Features: Traffic congestion patterns normally affect public holidays (Bao et al., 2017), and we included all public holidays in Ireland.

Traffic Congestion: TomTom is an online traffic index platform that collects GPS data about vehicle locations to determine traffic congestion [9]. It compares non-congested times with congested times, and the difference is quantified as a percentage expressing an increase in travel time. For example, a congestion level of 36% in Dublin means that a 30-minute trip will take 36% longer than the baseline.

3.2 Machine Learning Models

Predictions are often made using statistics and ML, one of the most popular types of computational intelligence (CI). Most statistical models are mathematically based, yet they struggle when faced with complex or nonlinear data. By contrast, ML can handle vast amounts of data, has a flexible modelling capability, can generalize and adapt, and are generally good at prediction tasks (Karlaftis and Vlahogianni, 2011). In the following, we review the ML models used in this study.

Multiple Linear Regression: The relationship between several independent variables and one dependent variable is explored with multiple linear regression, which is a variation of the ordinary least-squares regression model aiming to fit the linear relationship between explanatory variables.

Random Forest: Ensemble learning is an approach used to create a meta-model by combining a large number of weak models. The weak models are algorithms that are not suitable for learning complex models but are fast to learn and predict. RF (Breiman, 2001) is a bagging-based ensemble model that turns weak models (decision trees) into meta-models by making several copies of the training data and building decision trees based on the random subsets of features at each step. It is a model that could be used for linear and nonlinear relationships, and as it fits several decision trees to different sub-samples of the dataset and uses the average value, it allows accuracy to be improved and control over-fitting.

Light Gradient Boosting: Gradient Boosting (GB) is a boosting-based ensemble model that uses the original training data. To solve regression problems, GB first creates a single leaf

[6] https://data.smartdublin.ie/dataset/waste-data-thorntons-recycling

[7] https://data.smartdublin.ie/dataset/waste-data-aes-bord-na-mona

[8] https://www.google.com/covid19/mobility/

[9] https://github.com/ActiveConclusion/COVID19_mobility

Mean squared error	$\text{MSE} = \frac{1}{n}\sum_{t=1}^{n} e_t^2$
Root mean squared error	$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n} e_t^2}$
Mean absolute error	$\text{MAE} = \frac{1}{n}\sum_{t=1}^{n} \lvert e_t \rvert$
Mean absolute percentage error	$\text{MAPE} = \frac{100\%}{n}\sum_{t=1}^{n} \left\lvert \frac{e_t}{y_t} \right\rvert$
R-Squared	$\mathbb{R}^2 = 1 - \frac{SS_{res}}{SS_{tot}}$

Table 1. Evaluation metrics along with the formula.

representing the average of the target value for all samples and then adds it to a tree-based model that calculates residuals and then scales the tree's contribution to the final prediction by using the learning rate. This process continues depending on the number of trees and whether they improve the fit. GB models require tuning the tree depth, number of trees, and learning rate. One of the main issues of these GB algorithms is a high time-consuming computational cost for splitting the training datasets. In 2017, LGBM (Ke et al., 2017) solved this issue by introducing Gradient-Based One-Side Sampling (GOSS), which focuses on training data instances with more significant gradients, thereby contributing to the computation of information gain, also the Exclusive Feature Bundling (EFB) method that reduces the number of features by combining mutually exclusive features.

Support Vector Regression: SVR is an adaptation to SVM, which is used for regression analysis. SVR aims to find an n-dimensional hyperplane that can distinguishably classify data points and maximize the margin. SVR tries to fit the best line within a threshold value, the distance between the hyper-plane and boundary line, as opposed to other regression models that strive to minimize the difference between the actual value and the predicted value. SVR accuracy and computation complexity are heavily influenced by its hyper-parameters. Using the penalty parameter to punish samples whose errors exceed a specific value, the insensitive parameter to control the width of the insensitive zone to fit the training data, and the kernel function to handle both linear and nonlinear relationships are the main hyper-parameters.

3.3 Performance Evaluation

Table 1 shows the evaluation metrics used in this study. The error (e) shows the difference between estimated and actual values. MSE shows the average squared of the errors. RMSE is the square root of the MSE. MAE refers to how far the prediction differs from the actual value on average. MAPE indicates the accuracy of a forecasting method. R ^{2} is a statistic that shows the goodness of fit of a regression model, which is a comparison between the residual sum of squares (SS_{res}) and the total sum of squares (SS_{tot}).

3.4 Shapley Additive Explanations (SHAP)

Explainable AI sees ML as more than black boxes that produce interpretable results, which we use to interpret the output of ML models. Shapley (Lundberg et al., 2020) developed SHAP based on game theory by comparing the model's prediction with and without the feature in all ways possible. SHAP values can be utilized in a prediction task to analyze features that have the largest effect on model outcomes. Several studies (Zarbakhsh et al., 2022; Lundberg et al., 2020) applied SHAP techniques to further explain their features and contribution to their ML models. It is represented as a linear model where each SHAP value measures each sample's positive or negative contribution. SHAP has the advantage of being model-agnostic, which means that it could be calculated on any model, and each sample has its own set of SHAP values

Datasets	Size	Features
COVID-19 cases	1	Daily cases
COVID-19 vaccination	2	Daily vaccination Boosters vaccination
Mobility changes	6	Retail recreation Grocery pharmacy Parks Transit-stations Workplaces Residential
COVID-19 restrictions	8	School closing Work closing Cancel public events Restrictions on gatherings Close public transport Stay at home requirements Internal-movement restrictions International travel controls
Weather	2	Precipitation amount (mm) Minimum temperature (C) Maximum temperature (C)
Time	2	Holiday 1, Holiday 0
Street-waste	1	Waste quantity (Ton)
Traffic congestion	1	Traffic congestion changes (%) (Target Feature)

Table 2. Input and output features for ML algorithems.

4. EXPERIMENT

Fig. 2 illustrates our research framework for predicting traffic congestion in Dublin, which consists of four main steps: features construction, ML training, evaluation, and interpretation. We use the data sets explained in the section 3.1, which includes sets of categorical and numerical input features to predict traffic congestion values (output feature). Table 2 shows the data sets along with the size of each feature and their names, where all categorical features are converted to numerical values using one-hot encoding prepossessing. Furthermore, all the features are aggregated, and missing values are interpolated to make them suitable for building and training ML models. Following the cleansing and preparation of the data sets, we randomly split the data, using 80% for training and the remainder for testing. Following that, we trained five ML algorithms: MLR, RF, LGBM, and SVR, using the Radial basis function (RBF), known as the Gaussian kernel, which is the same as a Gaussian distribution function and Linear kernels. Given that SVR relies on distances between data points, we standardised features by removing the mean and scaling to unit variance to ensure that they fit into the same range. To improve the performance of our models, we used GridSearch CV and Random-Serch CV to tune the hyperparameters. The best-performing model was selected by five-fold cross-validation to estimate the accuracy of the ML model. Finally, to understand the model, the SHAP method was used to explain the contribution of the input features on traffic congestion prediction outcomes.

https://doi.org/10.5194/isprs-annals-X-4-W3-2022-301-2022 | © Author(s) 2022. CC BY 4.0 License.

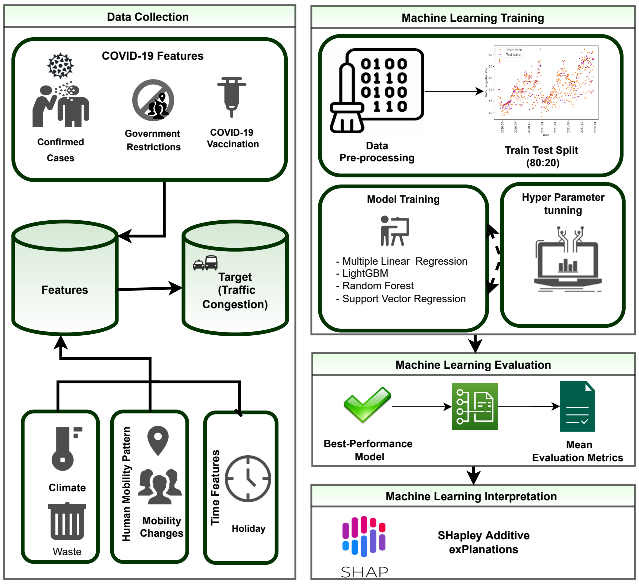


Figure 2. The research framework of the study.

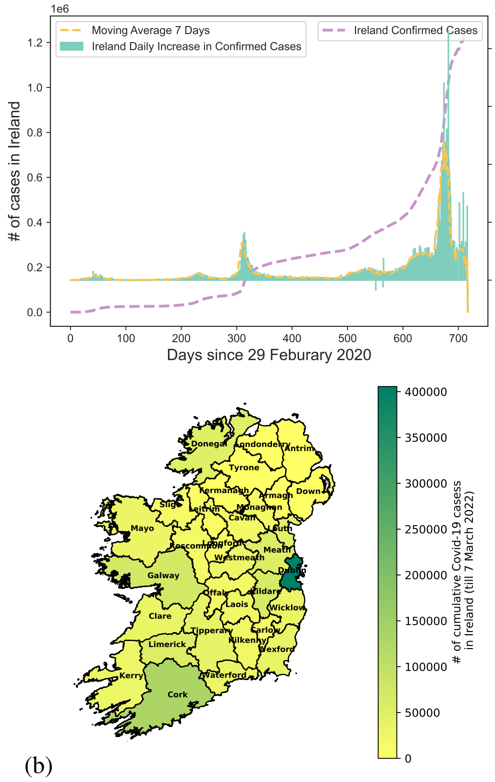
5. RESULTS

In this section, we start by analyzing the situation of COIVID-19 and its effect on travel behaviour and traffic congestion. Then we fit models and compare their results and interpret the best-performing model.

5.1 Travel Behaviour Changes During COVID-19

The COVID-19 pandemic has had a significant influence on travel behaviour due to government restrictions on movement and the transmission of disease; thus, to better understand these changes, we examine the COVID-19 situation and human mobility in Dublin in this section.
As of February 20, 2022, there were around 1.3 million confirmed cases of COVID-19 in the Republic of Ireland and more than 6000 fatalities. Fig 3 (a) shows the cumulative number of confirmed cases from February 29, 2020, along with the seven-day moving average and the daily number of cases. Two prominent peaks of confirmed cases occurred in Ireland during the Christmas-New Year period of 2021 and 2021, wherein the first spike, the government eased restrictions. In the second spike, the Omicron variant spread. Fig. 3 (b) displays the cumulative number of COVID-19 confirmed cases in all counties in Ireland. Dublin, Cork, and Galway, the main cities in Ireland, experienced the highest number of COVID-19 cases.

In response to such a high number of cases, the Irish government rolled out several NPIs to prevent diseases from spreading, Zarbakhsh et al. (2022) examined the impacts of government restrictions on human mobility and energy consumption patterns in Ireland and explained fluctuations in peak demand resulting from these changes in human behaviour. To further explain these changes in behaviour, Fig. 4 shows NPIs restrictions in conjunction with mobility changes in parks and the weight of biodegradable wastes collected from gardens and parks. Restriction severity is determined as the sum of the weighted average of eight government response measures in (Hale et al., 2021) Ireland using the stringency index approach described in, with a higher score indicating stricter government responses. The first confirmed case in Ireland was announced on February 29, 2020, which led to the closure of schools, colleges,



(a)

(b)

Figure 3. (a) The cumulative along with the daily number and seven-day moving average of COVID-19 cases in Ireland. (b) The cumulative number of confirmed cases in 26 counties of Ireland from the start of the pandemic till December 22, 2021.

and childcare facilities, followed by a stay-at-home order on March 27, 2020, indicating the highest government restrictions. The restrictions were eased from mid-May to late August, and 5-level restrictions were introduced in September. Early in November, the government issued Level 5 lockdown restrictions for the entire country, which gradually eased from December 1. In late December 2020, the third wave began, and the country moved to level-5 restrictions, bringing the line back to high severity. COVID-19 restrictions began to loosen throughout summer 2021 regardless of the Delta variant in June in Ireland. Due to the Omicron variant of COVID-19, some entertainment venues, bars, restaurants, and other businesses were urged to close early in December 2021. With these waves of COVID-19 in Ireland, park visitation patterns have changed. When severe restrictions like stay-at-home orders and level-5 restrictions are in place, mobility changes in parks are at their lowest level, leading to the lowest weight of biodegradable waste collected from gardens and parks. Comparing the histogram plots for April in 2020 and 2021, it can be seen that the high level of restrictions in the primary waves of Ireland has reduced the number of people visiting the parks; this, in turn, reduced the amount of waste collected from the parks. The street-waste in Dublin can also evidence such a change in mobility behaviour, which could be a result of less human mobility in Dublin's parks/gardens or recycling company employees working from home.

Dublin's pattern of traffic congestion variation shows an evident decline in traffic during public holidays, and there was a maximum of 22% traffic congestion in Dublin during the public holidays, meaning that travel times were on average 6.6 minutes longer than the baseline non-congested conditions. However,

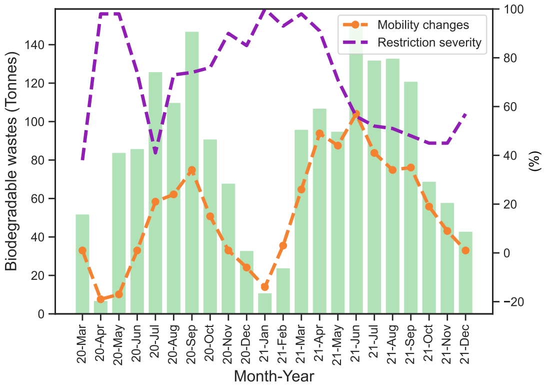


Figure 4. Monthly net weight (Ton) of biodegradable waste collected from gardens/parks from March 2020 to December 2021, along with restrictions and mobility changes in parks.

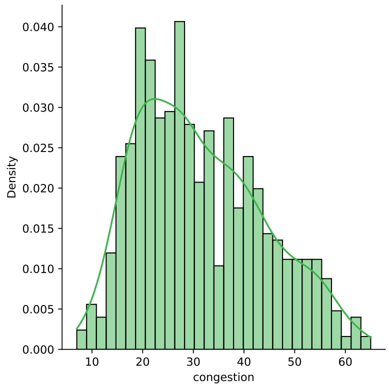


Figure 5. A density plot of traffic congestion in Dublin using a kernel density estimation of the distribution.

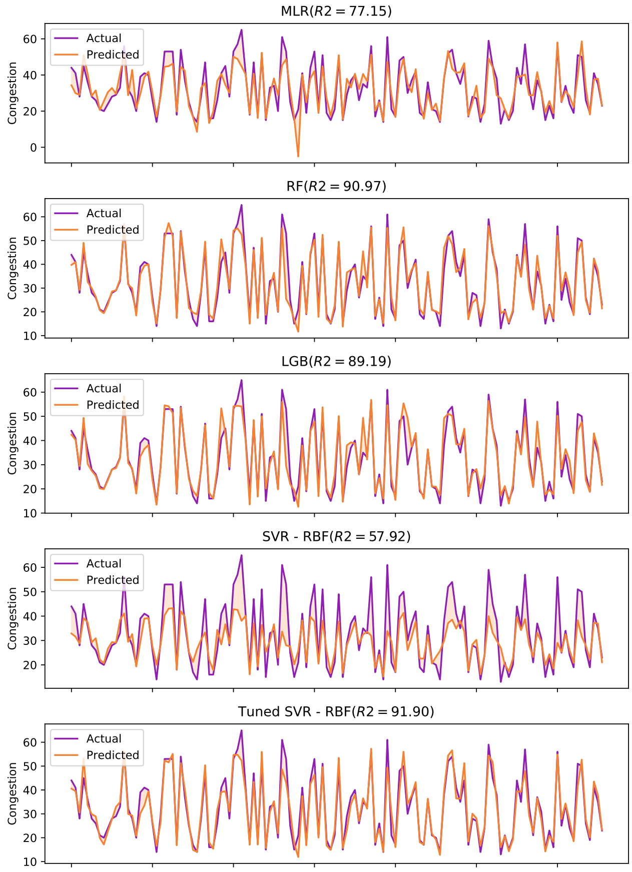


Figure 6. The comparison between model prediction and actual value. The five plots show the actual value of traffic congestion versus the predicted value over time by reporting the $\mathbb{R}^2$.

the average extra travel time reached around 20.2 minutes on non-holiday days during the COVID-19 pandemic period. To better predict traffic congestion during COIVD-19, we consider the explained features as explanatory features.

5.2 ML Fitting and Evaluation

Target Feature: Traffic congestion in this study is displayed in percentage, which contrasts traffic congestion with baseline non-congested conditions using the TomTom traffic index, which could also convert to traffic time. Traffic congestion in Dublin during the COVID-19 ranges from 7% to 65%, with a mean value of 31.4%. Fig. 5 displays a density plot that illustrates the distribution of traffic congestion over a continuous interval. Dublin traffic congestion has a standard deviation of 12.4, with a skew of 0.48 and a Kurtosis of -0.53, showing a reasonably symmetrical distribution.

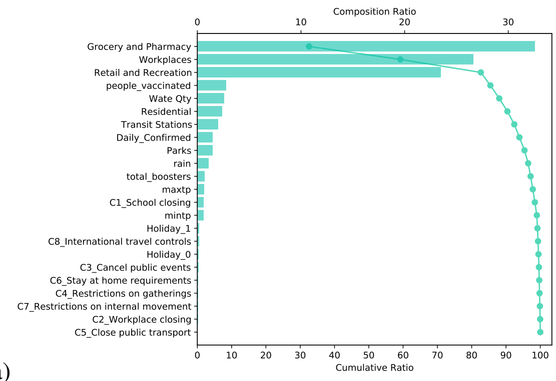
ML Training: After training MLR, RF, LGBM and SVR models on training data, we tested the model on test sets. Unlike linear regression, which minimizes the difference between actual and predicted values through the best-fit line, SVR finds the best line by fitting it within an epsilon threshold and can estimate the nonlinear function. Hyperparameter tuning is essential for SVR algorithms, and we improved the model prediction significantly by estimating the best hyperparameter using GridSearchCV. *Gamma* is defined as the degree of non-linearity; *epsilon* supports finding a margin of tolerance tube, in which only objects lying outside the tube around the estimated hyperplane are penalized (Carrasco et al., 2019), and *C* is defined as a regularization parameter. with hyper-parameter of kernel= RBF, gamma = 0.01, epsilon = 0.5 and C = 1000 SVR model retrained and the results are reported.

ML Evaluation: Table 3 shows the evaluation metric for all models based on the random train/test split. Based on a comparison of the four evaluation metrics, SVR and RF have the best performance in terms of the lowest MSE, RMSE, MAE, MAPE, and highest $\mathbb{R}^2$ values. The results of LGBM are almost as good as those of RF. MLR, however, shows more significant errors. Fig. 6 compares the actual value with the predicted value for all models and reports the $\mathbb{R}^2$. To ensure that the best-performing models, RF and SVR (RBF), do not result from selecting the best part of the data randomly, we applied a five-fold cross-validation technique to train and test our model on five sets of data. RF and SVR (RBF) models are trained and tested five times; for every iteration, one fold is only used as a test set and the rest as training data. The calculation of the evaluation metrics yield to MSE of 12.28, RMSE of 3.49, MAE of 2.42, MAPE of 8.6%, and $\mathbb{R}^2$ of 92.0%, surpassing the results of the random-split as the average of different folds reported for the RF model. For the SVR model, MSE of 14.99, RMSE of 3.86, MAE of 2.64, MAPE of 8.6%, and $\mathbb{R}^2$ of 90% is slightly less than the random-split. In the following, we choose RF as a best-performing model and interpret the feature contributions.

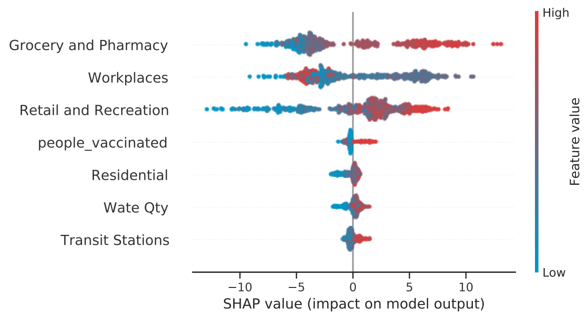
ML Interpretation: Fig. 7 (a) shows the most significant features affecting the performance of the RF model. It displays the global importance of features defined by the average absolute Shapley value per feature, with all features sorted according to this metric. It is evident that during COVID-19, mobil-

Models/ Metric	MLR	RF	LGBM	SVR	Tuned-SVR
MSE	43.1	17.0	20.4	79.5	15.3
RMSE	6.5	4.1	4.5	8.9	3.9
MAE	4.9	2.5	2.8	6.1	2.8
MAPE	16.1	8.6	9.3	18	8.6

Table 3. Comparing the evaluation metrics for several ML algorithms based on the random-split.



(a)



(b)

Figure 7. (a) RF SHAP waterfall plot: It depicts the most influential features by sorting all features by the SHAP absolute mean values along with the cumulative ratio of their importance. (b) RF SHAP summary plot: Features are sorted according to the sum of SHAP value magnitudes over all samples. The color represents the features value (red high, blue low).

ity changes in three categories of grocery and pharmacy, workplace, retail and recreation, vaccination rate, and the amount of waste on the streets were the essential features of the model, and they accounted for almost 86% of RF interpretations. As shown by the green line, the mobility changes of groceries and pharmacies contribute around 33% of the interpretability of the model, and the mobility changes of workplaces contribute around 27%.

Fig. 7 (b) depicts a SHAP summary plot of the first seven influential features in which features are arranged according to their importance (mean absolute SHAP values) in predicting traffic congestion. The horizontal axis displays the SHAP values for each feature, and a deeper red colour indicates a higher value of the feature, while blue colours indicate a lower value of the feature that affected the RF model outcome. This reveals that a high value of human mobility in grocery and pharmacy, retail and recreation, and transit stations increases the predicted traffic congestion in Dublin city for the RF model. This trend is also could be observed in the street-waste feature.
Results suggest that the amount of street-waste and human mobility-related features could be a strong proxy for predicting traffic congestion in an unprecedented situation like COVID-19.

6. CONCLUSION

This study explores the potential of novel mobility-based features and urban street-waste data to predict traffic congestion during COVID-19. We train several ML algorithms to predict traffic congestion in Dublin, enhance their performance, and propose the RF model as the best-performing model. We explain this model further by identifying the most significant features contributing to the model outcomes. According to the SHAP calculation, mobility changes in grocery and pharmacy, workplace, retail, and recreation, as well as vaccination rates and urban street-waste, are the best predictors. This study indicates that during unprecedented events like COVID-19, the effects of government restrictions and human mobility are visible in human behavior, and the amount of urban street-waste indicates the street movement patterns and has the potential to describe traffic dynamics.

This study can assist urban planners in designing a smart city with a resilient, intelligent transport system and suggest considering human mobility changes in different categorical places along with street-waste collected from cities as a novel feature to predict traffic congestion in cities. It could also help authorities understand the effects of their precautions measures during crises on traffic congestion. Further extensions of this study is expanding the current study to more countries, considering the spatiotemporal traffic dynamics patterns in different cities, and comparing the human mobility-based features on traffic congestion during COVID-19.

REFERENCES

Bao, Y., Xiao, F., Gao, Z., Gao, Z., 2017. Investigation of the traffic congestion during public holiday and the impact of the toll-exemption policy. *Transportation Research Part B: Methodological*, 104, 58–81.

Bratsas, C., Koupidis, K., Salanova, J.-M., Giannakopoulos, K., Kaloudis, A., Aifadopoulou, G., 2020. A comparison of machine learning methods for the prediction of traffic speed in urban places. *Sustainability*, 12(1), 142.

Breiman, L., 2001. Random forests. *Machine learning*, 45(1), 5–32.

Carrasco, M., López, J., Maldonado, S., 2019. Epsilon-nonparallel support vector regression. *Applied Intelligence*, 49(12), 4223–4236.

Chen, C., Zhao, X., Liu, H., Ren, G., Zhang, Y., Liu, X., 2019. Assessing the influence of adverse weather on traffic flow characteristics using a driving simulator and VISSIM. *Sustainability*, 11(3), 830.

Hale, T., Angrist, N., Goldszmidt, R., Kira, B., Petherick, A., Phillips, T., Webster, S., Cameron-Blake, E., Hallas, L., Majumdar, S. et al., 2021. A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker). *Nature Human Behaviour*, 5(4), 529–538.

Hara, Y., Kuwahara, M., 2015. Traffic Monitoring immediately after a major natural disaster as revealed by probe data–A case in Ishinomaki after the Great East Japan Earthquake. *Transportation research part A: policy and practice*, 75, 1–15.

Karlaftis, M. G., Vlahogianni, E. I., 2011. Statistical methods versus neural networks in transportation research: Differences, similarities and some insights. *Transportation Research Part C: Emerging Technologies*, 19(3), 387–399.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.-Y., 2017. Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, 30, 3146–3154.

Kraemer, M. U., Yang, C.-H., Gutierrez, B., Wu, C.-H., Klein, B., Pigott, D. M., Du Plessis, L., Faria, N. R., Li, R., Hanage, W. P. et al., 2020. The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science*, 368(6490), 493–497.

Li, H., Zhang, Y., Zhu, M., Ren, G., 2021a. Impacts of COVID-19 on the usage of public bicycle share in London. *Transportation Research Part A: Policy and Practice*, 150, 140–155.

Li, J., Xu, P., Li, W., 2021b. Urban road congestion patterns under the COVID-19 pandemic: A case study in Shanghai. *International Journal of Transportation Science and Technology*, 10(2), 212–222.

Li, Q., Bessell, L., Xiao, X., Fan, C., Gao, X., Mostafavi, A., 2021c. Disparate patterns of movements and visits to points of interest located in urban hotspots across US metropolitan cities during COVID-19. *Royal Society open science*, 8(1), 201209.

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., Lee, S.-I., 2020. From local explanations to global understanding with explainable AI for trees. *Nature machine intelligence*, 2(1), 56–67.

Ma, X., Yu, H., Wang, Y., Wang, Y., 2015. Large-scale transportation network congestion evolution prediction using deep learning theory. *PloS one*, 10(3), e0119044.

Mridha, S., Ghosh, S., Singh, R., Bhattacharya, S., Ganguly, N., 2017. Mining twitter and taxi data for predicting taxi pickup hotspots. *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017*, 27–30.

Pan, B., Zheng, Y., Wilkie, D., Shahabi, C., 2013. Crowd sensing of traffic anomalies based on human mobility and social media. *Proceedings of the 21st ACM SIGSPATIAL international conference on advances in geographic information systems*, 344–353.

Qiao, Y., Si, Z., Zhang, Y., Abdesslem, F. B., Zhang, X., Yang, J., 2018. A hybrid Markov-based model for human mobility prediction. *Neurocomputing*, 278, 99–109.

Qiu, B., Fan, W., 2021. Travel time forecasting on a freeway corridor: a dynamic information fusion model based on the random forests approach. *Smart and Resilient Transport*.

Song, X., Kanasugi, H., Shibasaki, R., 2016. Deeptransport: Prediction and simulation of human mobility and transportation mode at a citywide level. *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, 2618–2624.

Sung, J., Monschauer, Y., 2020. Changes in transport behaviour during the Covid-19 crisis. Data retrieved from `https://www.iea.org/articles/changes-in-transport-behaviour-during-the-covid-19-crisis` (21 March 2022, date last accessed).

Taleb, N. N., 2007. *The black swan: The impact of the highly improbable*. 2, Random house.

Toan, T. D., Truong, V.-H., 2021. Support vector machine for short-term traffic flow prediction and improvement of its model training using nearest neighbor approach. *Transportation research record*, 2675(4), 362–373.

TomTom, 2021. Tomtom traffic index. Data retrieved from `https://github.com/ActiveConclusion/COVID19_mobility` (21 February 2022, date last accessed).

Wang, W., Miao, W., Liu, Y., Deng, Y., Cao, Y., 2022. The impact of COVID-19 on the ride-sharing industry and its recovery: Causal evidence from China. *Transportation Research Part A: Policy and Practice*, 155, 128–141.

Wolshon, B., McArdle, B., 2009. Temporospatial analysis of Hurricane Katrina regional evacuation traffic patterns. *Journal of Infrastructure Systems*, 15(1), 12–20.

Wu, F., Wang, H., Li, Z., 2016. Interpreting traffic dynamics using ubiquitous urban data. *Proceedings of the 24th ACM SIGSPATIAL international conference on advances in geographic information systems*, 1–4.

Xie, Y., Zhang, Y., Ye, Z., 2007. Short-term traffic volume forecasting using Kalman filter with discrete wavelet decomposition. *Computer-Aided Civil and Infrastructure Engineering*, 22(5), 326–334.

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

Yu, G., Zhang, C., 2004. Switching arima model based forecasting for traffic flow. *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2, IEEE, ii–429.

Yu, H., Wu, Z., Wang, S., Wang, Y., Ma, X., 2017. Spatiotemporal recurrent convolutional networks for traffic prediction in transportation networks. *Sensors*, 17(7), 1501.

Zarbakhsh, N., Saeed Misaghian, M., McArdle, G., 2022. Human Mobility-Based Features to Analyse the Impact of COVID-19 on Power System Operation of Ireland. *IEEE Open Access Journal of Power and Energy*, 1-1.

Zhang, L., Alharbe, N. R., Luo, G., Yao, Z., Li, Y., 2018. A hybrid forecasting framework based on support vector regression with a modified genetic algorithm and a random forest for traffic flow prediction. *Tsinghua Science and Technology*, 23(4), 479–492.