

Received 5 April 2023, accepted 30 April 2023, date of publication 3 May 2023, date of current version 9 May 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3272662

RESEARCH ARTICLE

Machine Learning for Transport Policy Interventions on Air Quality

FARZANEH FARHADI¹, ROBERTO PALACIN, AND PHIL BLYTHE

School of Engineering, Newcastle University, NE1 7RU Newcastle upon Tyne, U.K.

Corresponding author: Farzaneh Farhadi (F.Farhadi2@newcastle.ac.uk)

This work was supported in part by the Engineering and Physical Sciences Research Council (EPSRC) under Grant EP/V519571/1, and in part by the Arup Group Ltd.

ABSTRACT Air pollution reduction is a major objective for transport policy makers. This paper considers interventions in the form of clean air zones, and provide a machine learning approach to assess whether the objectives of the policy are achieved under the designed intervention. The dataset from the Newcastle Urban Observatory is used. The paper first tackles the challenge of finding datasets that are relevant to the policy objective. Focusing on the reduction of nitrogen dioxide (NO₂) concentrations, different machine learning algorithms are used to build models. The paper then addresses the challenge of validating the policy objective by comparing the NO₂ concentrations of the zone in the two cases of with and without the intervention. A recurrent neural network is developed that can successfully predict the NO₂ concentration with root mean square error of 0.95.

INDEX TERMS Air quality, clean air zone, data-driven framework, machine learning, policy validation, transportation system.

I. INTRODUCTION

Clean air zones are being designed and implemented by local authorities to improve the air quality. The clean air zone can apply specific requirements to both commercial operators and private motorists. This include buses, taxis, heavy good vehicles, and private cars that are not compliant with the intended emission standards. The UK government's Clean Air Strategy [1] includes the implementation of clean air zones in large UK cities. The main purpose of introducing and implementing clean air zones is to reduce the air pollution levels and satisfy the legal requirements on keeping specific pollutants (e.g., nitrogen oxides) below the allowable limits. The pollutants include nitrogen oxides (NO_x), particulate matters (PM_x), carbon dioxide (CO₂), and other greenhouse gases. In particular, breathing air that has high concentrations of NO₂ can create irritation in the human respiratory system, and can cause diseases such as asthma [2]. Once properly designed and implemented, a clean air zone reduces traffic emissions that are harmful, improves air quality and protects people's health (see the work [3] on the performance of such

zones in European cities). The zone can affect the users' behaviours by putting appropriate changes. Designing clean air zone interventions and increasing their effectiveness are essential for achieving an improved air quality, and data will play a central role in such a design.

The recent advances in sensing and the storage of large volumes of data have introduced a revolution in the way we design and manage transportation systems [4], [5]. Real-time and historical information gathered from the transportation network enable us to learn, and develop interventions to improve the efficiency of the network and make it sustainable. This has opened new research directions in the transport research community and has generated a strong interest in the relevant industries and among policy makers to move towards intelligent transportation systems in which data collection and analysis plays a fundamental role in decision making, design, and increased efficiency [6], [7], [8]. The survey papers [4] and [5] analyse the latest research efforts on developing data-driven intelligent transport systems while discussing the functionality of the key components and future research directions to tackle the related challenges.

Using large volume of data effectively in the transport sector bring its own challenges and opportunities for further

The associate editor coordinating the review of this manuscript and approving it for publication was Vicente Alarcon-Aquino¹.

research. Specific data-driven solutions that have been studied by researchers recently to address challenges in the transport sector are as follows. The work [9] studies automation of operations in seaport logistics and proposes a Big Data architecture for secure data sharing and promoting an intelligent transport multimodal terminal for improving decision making. The work [10] proposes a data-driven approach to construct an accurate model for predicting short-term traffic flow by combining the spatio-temporal analysis with a Gated Recurrent Unit. The article [11] proposes a convolutional neural network architecture for predicting multi-lane short-term traffic flow. Other applications of data-driven methods in the transport sector include building preventive maintenance decision models of urban transportation systems [12], optimising fuel consumption and sulfur oxide (SO_x) emissions using big data analytics techniques to make environmentally sustainable operations in maritime shipping [13], and predicting transportation carbon emission using urban big data to mitigate climate change [14], [15]. Unfortunately, the previous research on validating objectives of transport policy interventions using data-driven methods is very limited.

The main contribution of this paper is to provide data-driven frameworks for identifying the relevant data types and for using machine learning techniques to check whether the objectives of a policy will be achieved. We apply the frameworks to the use case of a clean air zone where the objective is to improve air quality in areas where there are currently NO₂ exceedances. The intervention implemented in the clean air zone is considered to be in the form of specific charges on vehicles entering the zone dependent on their emission levels. We use the datasets from the Newcastle Urban Observatory [16]. The aims of this paper are as follows: (a) Identify, gather, integrate, preprocess and analyse data types relevant to the policy objective (which could be for instance in the form of reducing NO₂ concentrations, or reducing the number of days with high NO₂ concentrations); (b) Develop suitable machine learning methods based on the input processed data and validate their outputs with respect to the clean air zone objectives; and (c) Make predictions using the learned models to check that the policy objectives will be achieved.

A small subset of the results of this paper is accepted for presentation at the ITSC conference [17]. This journal paper substantially extends the results of the conference paper in multiple directions. First and foremost, we provide a framework for finding the relevant data types in addition to a framework for validating the intervention while [17] only provide one framework. Second, we have applied our framework using four different machine learning models in addition to the time series analysis provided also in [17]. We have also completed our paper with an extensive literature review and reported different performance metrics of the learned models.

This paper is organised into the following sections. Section II gives an overview of the Newcastle Urban Observatory and the policy objective considered in this work. Sections III and IV review the related work and machine

learning techniques used in air quality. The method of this research is presented in Section V. A brief overview of the dataset and the policy objective is provided. We also present our frameworks designed to tackle the challenges that are the focus of this paper: a framework for finding datasets that are related to the policy objective, and a framework for validating the policy objective targeted at air quality. In Section VI, we provide the details of the machine learning models used in our study (Decision Trees, Light Gradient Boosting Machine, K-Nearest Neighbours, and Gradient Boosted Decision Tree). We also give the test metrics that can be used to assess the performance of the learned model. These metrics include *accuracy*, *precision*, *recall*, *F1 score*, and *confusion matrix*, and will be defined in this section. We also discuss the details of another main contribution of the paper, which is a time series analysis of the data for making predictions on the air quality. Section VII presents the results and discussions on applying the frameworks on the dataset from Urban Observatory at Newcastle, United Kingdom. Finally, we conclude the paper in Section VIII with a summary and our future research directions.

II. THE NEWCASTLE URBAN OBSERVATORY

In this paper, we will use the clean air zone intervention as a case study and will apply the results using datasets from Newcastle upon Tyne as the City Council is planning to design and implement a clean air zone to be open from January 2023. We employ data available from Newcastle Urban Observatory [16] that initially funded under the UK Collaboratorium for Research in Infrastructure and Cities (UKCRIC).¹

Newcastle upon Tyne is the largest city in the North East of England with approximate population of 300,000. The Newcastle Urban Observatory (UO) was developed with a multi-million-pound investment to serve as a large-scale data capturing infrastructure. The open dataset of UO is used by community groups, local and national government, and research projects ranging from cyber-security to quantifying the impact of COVID measure and flood forecasting. The data handled by the UO covers a wide range of city metrics including mobility, air quality, climatic variables, and infrastructure.

The volume of data in UO is in the order of billions of data points that are published as anonymous open data [18]. The dataset includes 900 million data points measured since 2016, 60 data types, and 2000 observations every minute. Figure 1 shows the geographical locations of sensors that measure and send data to the UO. The majority of the sensors are located in Newcastle upon Tyne and its surrounding areas. The map shows clusters of sensors for a better visualisation. Note that the measurements stored in the UO are raw data and needs to be processed for improving the quality of the data. We have used a subset of the measurements and preprocessed the dataset as will described in Section V-C.

¹<https://urbanobservatory.ac.uk/explore/ukcric>

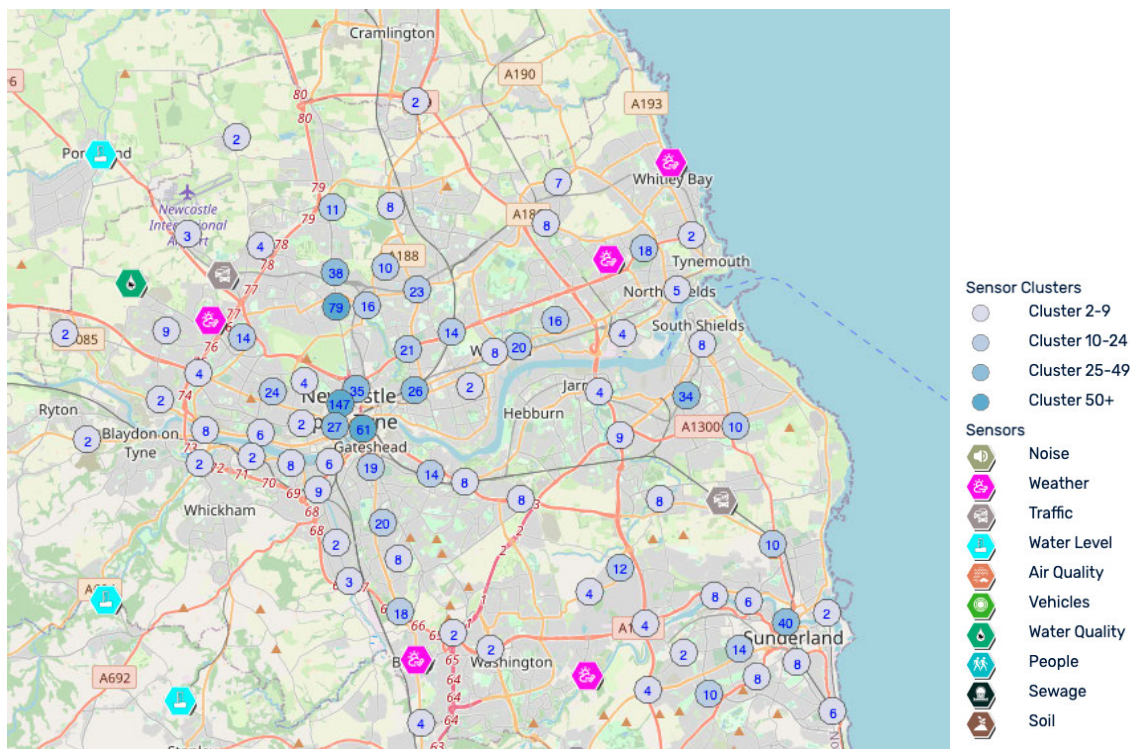


FIGURE 1. Geographical locations of sensors that measure and send data to the Urban Observatory. Circles with numbers show clusters of sensors.

The data being monitored and stored at the Urban observatory sites include different themes and each theme has different variables. The themes considered for our case study include *air quality*, *weather*, *traffic* and *timestamp*. There are different sensors for monitoring these themes. These sensors are installed in various locations of Newcastle. The sensors that we have considered are close to Newcastle City Centre to have a better understanding of the air quality at the centre of Newcastle. Each theme has different variables. The variables for the *air quality* theme that we consider include CO, PM_{2.5}, PM₁, PM₁₀, PM₄, O₃, NO, and NO₂. The variables that we consider for *weather* theme include Wind Direction, Wind Speed, Solar Radiation, Solar Diffuse Radiation, Pressure, Rain Duration, Rain ACC and Max Wind Speed. The variables for the *traffic* theme that we consider include Traffic Flow and Average Speed, which are both collected by the Newcastle UO and by the Urban Traffic Management and Control at a greater fidelity. We also have *timestamp* theme that include *Year*, *Month*, *Day*, *Hour*, *Minute*, and *Second*. This means the measurements are taken in general every second and stored in the database of the Newcastle Urban Observatory.

A. POLICY OBJECTIVE AND THE INTERVENTION

The cabinet members at Newcastle and Gateshead Councils have confirmed the plans for introducing a clean air zone to operate in Newcastle city centre [19]. The planned date for introducing the intervention is January 2023. The zone will include the city centre of Newcastle and routes over the Tyne,

High Level, Swing and Redheugh bridges. The intervention will impose charges on all buses, taxis, coaches, vans and heavy goods vehicles (HGVs) that do not meet the emissions standards of EUROIV for petrol and EUROVI for diesel vehicles. The primary goal of the clean air zone in Newcastle is to improve the poor air quality. Therefore, we consider the following as the objective of the introduced policy:

After introducing the clean air zone at Newcastle, the concentration of NO₂ will reduce. More specifically, the time duration when the concentration of NO₂ is unhealthy (NO₂ concentration above 100 parts per billion) will be reduced by 10%.

Note that we have selected 10% reduction as an example for a proof of concept to demonstrate the usefulness of the frameworks designed in this paper. This reduction can be estimated by the Clean Air Zone experts in their technical documents backed up by air quality modelling, operational cost modelling, and behavioural response estimates.²

We have designed two frameworks to tackle the specific challenges of validating policy objectives using large datasets. The first framework addresses the challenge of finding data types that are related to the policy objective using machine learning techniques. The second framework validates an intervention and checks how well the objectives of the policy are achieved. These framework will be presented in

²<https://cleanairgm.com/technical-documents>

Section V after reviewing the related literature on air quality and machine learning in Sections III–IV.

III. RELATED WORK

In the United Kingdom, the Department for Transport (DfT) and the Department for Environment, Food & Rural Affairs (Defra) have jointly produced a report that lays out a framework for the design and operation of clean air zones in England [20]. It recommends the approach to be taken by local authorities when implementing and operating a clean air zone. These recommendations apply also to the clean air zone being considered for implementation in Newcastle. An example of such an implementation in England is the Greater Manchester Clean Air Zone (GMCAZ) [21]. The interventions designed in the GMCAZ was launched on 30 May 2022 and requires the vehicles that do not meet the emissions standards, to pay a fee when entering the clean air zone. The non-compliant vehicles are: Heavy goods vehicles, buses, coaches, vans, minibuses, private hire cars, and motor-caravans that have a EUROIV or earlier diesel engine that have a EUROIII or earlier petrol engine. Private cars are not currently affected by the intervention. This new scheme also provides financial support with more than £120m of government funding to help businesses in the region, organisations and people switch to compliant vehicles by either replacing or retrofitting non-compliant vehicles. The technical reports published in the website of GMCAZ³ clearly show the role of data in designing the related interventions and evaluating them when they are implemented. In general, datasets play two main roles: (1) datasets are used to select and tune parameters of the physical models developed for air quality, and (2) datasets are used for monitoring and checking if the policy objectives are achieved (e.g., reducing the pollutant level to some value).

In contrast to a clean air zone in which the local authority is actively trying to improve the quality of the air, *low emission zones* put restrictions specifically on vehicles that do not meet a minimum standard for vehicle emissions, e.g., the European Union's emissions standards [22] on harmful air pollutants and greenhouse gases. The paper [23] studied how the low emission zone implemented in London impacted the vehicle usage and air pollution by using the data from the registration and enforcement information. The authors focus on concentrations of particulate matter PM_{10} and NO_x . This choice was made due to the fact that in London, approximately 25% of PM_{10} and 57% of NO_x emissions from road transport come from heavy vehicles [24]. Using ambient air quality measurements, they showed that concentrations of particulate matter have dropped by 2.5–3.1% within the low emission zone, but they did not find any noticeable differences for all measured NO_x concentrations.

Another line of research studies various methods for specifying and deciding on boundaries of a clean air zone. The paper [25] uses data from the air quality monitoring stations

and combines statistical analyses with interpolation techniques to identify the areas with the highest concentrations of particulate matter. The paper [26] integrates an atmospheric model with a kinematic model to identify the boundaries of the catchment of air affecting concentrations of air pollution. The available data is used to initialise and calibrate the models. The paper [27] discusses an empirical approach called *participatory modelling* to find spatial representations of local knowledge about air pollution and use them in local governance of air quality. They present empirical data from a three-city case study and generate maps using the local knowledge, which can then be used as a form of consultation for local governance of the politics on air pollution.

Although the above works show the potential of data-driven techniques for modelling, analysing, and addressing challenges related to air quality, developing data-driven methods for validating policy objectives on air quality has not received a considerable attention. In the rest of this paper, we review machine learning models used to solve problems on air quality and show how to develop such models for the clean air zone of Newcastle using the available data.

IV. MACHINE LEARNING AND AIR QUALITY

Data plays a central role in *smart cities*, which are cities that integrate information, communication, and computing technologies with citizens and critical infrastructure components and services of the city to facilitate sustainable development and improve the life quality [28], [29], [30]. In the context of smart cities, applying machine learning methods for air quality prediction has recently gained attention from researchers. A systematic review of data-driven methods for air pollution prediction is provided in [29]. Their work provides an overview of machine learning techniques used in the smart city domain to predict air quality, and classifies the temporal resolutions analysed with these techniques. Prediction of NO_2 concentration in an air quality monitoring site of the Greater Manchester Area (UK) is performed in [31] using a statistical model called ARIMAX. Their results show that the performance of ARIMAX is similar to standard statistical approaches in terms of the difference between simulated and measured concentrations, however the accuracy of ARIMAX in the prediction of extreme air pollution events is 27% better than the standard statistical methods and 113% better than using neural network models. The paper [32] has focused on air quality prediction in Madrid, Spain. This work provides a method to predict CO, NO_2 , O_3 , PM_{10} , SO_2 and pollen concentrations using long short term memory (LSTM). It tries to find the best configuration in the LSTM (e.g., how neurons are connected) for reducing the prediction error and having robust one day-ahead air quality predictions. The work does not include any results on using these predictions for intervention validation or design.

The paper [33] considers the travel restrictions and the lockdowns imposed due to the COVID-19 pandemic, and studies their impact on Air Quality using machine learning methods. The work uses Gradient Boosting Machine

³<https://cleanaigm.com/technical-documents>

algorithm to assess the impact of full or partial lockdown on air quality in Quito, Ecuador. The approach in [33] is to use pre-lockdown data to predict the pollution levels without lockdown and then compare the predictions with pollution levels measured under lockdown measures. Another study is performed in [34] to understand the effect of COVID-19 restrictions on CO₂ emissions. It uses CO₂ observations and an atmospheric transport model to compute changes in CO₂ emissions caused by the imposed lockdown. It predicts a 30% decrease in CO₂ emissions and concludes that this reduction is mainly due to changes in road vehicles as opposed to other non-traffic emissions which showed small changes.

The paper [35] applies support vector regression (SVR) as a machine learning approach to predict the air quality index (AQI). AQI is an index for quantifying the level of pollution of air. Its values range from 0 to 500 parts per billion (ppb), where higher values indicate larger pollutions. The SVR model used in [35] is a nonlinear mapping that maps the dataset into a feature space and fits a linear regression model to the dataset in the new feature space. A kernel function is defined as a mapping from the input space to the new feature space. The paper shows empirically that radial basis function as kernel functions gives the best result for prediction on the chosen dataset. The approach is implemented on a dataset containing hourly data measured from the state of California, USA, between January 1, 2016, and May 1, 2018. The result shows that the pollutant concentrations can be successfully predicted to using SVR method as a regression problem, and the six categories of AQI can be predicted as a classification problem with an accuracy of 94.1%. The paper does not study the effect of interventions for improving air quality.

The paper [36] uses three machine learning methods for predicting concentrations of PM₁₀ and PM_{2.5} using road traffic, meteorological data and pollutant data measured and stored at Air Quality Monitoring sites of London. The machine learning models used in the study of [36] include Artificial Neural Networks (ANN), Boosted Regression Trees (BRT) and Support Vector Machines (SVM). The reported implementations show that ANN and BRT are better than SVM in predicting PM₁₀ and PM_{2.5} concentrations and these two models can be applied in managing the traffic-related particulate matter concentrations. The authors also conceptualised a hypothetical scenario to demonstrate the use of machine learning models in air quality management. The scenario assumed that the study area permits only EUROIV petrol and EUROVI diesel vehicles to be driven in that area. The dataset is the revised using the Emissions Factors Toolkit (EFT) [37] and new machine learning models are constructed on the modified data. The paper demonstrate that machine learning methods can be used to forecast concentrations of pollutants PM₁₀ and PM_{2.5} whenever rich datasets are available.

The paper [38] has developed a Bayesian network method with an optimised configuration to provide a probabilistic

traffic data analysis and to predict traffic-related air pollutions. Machine learning predictive models are developed in [39] for predicting particulate matter concentration using Taiwan Air Quality Monitoring datasets from 2012 to 2017. The developed predictive models were compared with the traditional models and cross-validation is used to select the best model with the highest performance. The paper [40] has studied reduction of ambient air pollution and congestion using weather forecasts and predictive cordon tolls. The authors used a model of emission dispersion to forecast air quality using recorded weather data for Tehran in 2016. It is shown that the constructed pricing scheme decreases the daily average CO concentrations.

Although the works reviewed above make promising observations on the use of machine learning methods for making predictions on air quality, these works do not give a framework for efficient analysis of policy interventions related to air quality. In the next sections, we provide a general data-driven framework for analysis of the policy objectives that has machine learning methods as a core modelling and computation component. As a proof of concept, we study the reduction in concentration of NO₂ by the implemented clean air zone.

V. METHODS AND DATA

In this section, we present two frameworks for identifying data types that are the most relevant to a policy objective, and for checking how well the objectives of a policy are achieved. We use the term ‘framework’ since our approaches presented in this section are high level and flexible, and can be applied to different policy objectives. The details and the choice of machine learning models can be decided depending on the specifics of the policy objective.

A. IDENTIFYING RELEVANT DATA TYPES

We have designed our first framework for identifying data types that are the most relevant to a policy objective. The rationale behind designing our framework is to use well-established machine learning methods that do not require an understanding of air pollutants’ physical or chemical properties but need sufficiently rich datasets. This framework sets the first steps we need to take in order to capture the complex nonlinear relationships between the concentration of air pollutants and meteorological variables in machine learning models.

The framework is presented in Figure 2. The first step in this framework is to analyse a given policy objective to extract the variables that are important in assessing the successful implementation of the intervention. These variables are the ones mentioned explicitly in the intervention documents while specifying quantitatively how much they are expected to change after the implementation of the intervention. Once these target variables are identified, the next step is to analyse the data from these variables using machine learning methods to measure their relative importance in predicting and

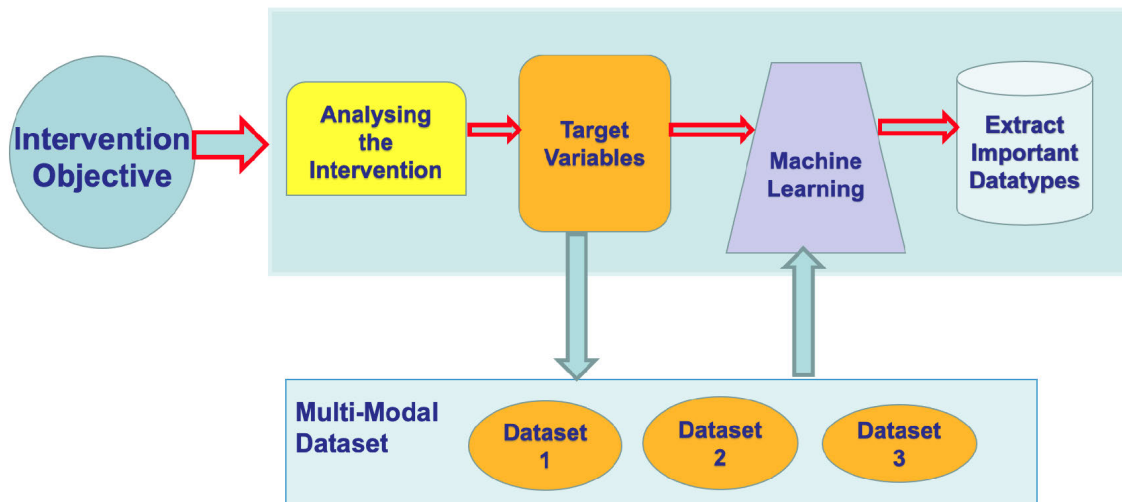


FIGURE 2. A framework for identifying data types that are relevant to an intervention.

affecting the objective of the policy. This relative importance can then be used to select a subset of datasets that make accurate predictions for assessing the policy objective. In order to compute the relative importance of the data types and extract the important ones, we use two different data-driven methods. The first one is based on *Pearson correlation coefficient*. The correlation takes values in the interval $[-1, 1]$, a value of ± 1 indicates that two variables are dependent linearly, and values closer to 0 means they become independent of each other. A plus sign shows positive relationship and a minus sign shows a negative relationship. The second one is based on *permutation feature importance*, which computes the importance of a feature by first training the model on the train dataset, then permuting the feature and computing the increase in the prediction error of the model. If a feature is important in making predictions, the prediction error should increase after permutation. If a feature is unimportant, the change in the prediction error will be by doing the permutation. In general, feature importance can take any values. We normalise the feature importance to obtain values in the range $[0, 1]$, which show the relative importance of the features.

B. INTERVENTION VALIDATING FRAMEWORK

We have designed a second framework for validating an intervention, and checking how well the objectives of an intervention are achieved. This framework is presented in Figure 3. The underlying idea of this framework is to compare the behaviour of the transportation system under study in the two cases of with and without intervention. Historical data and new data gathered after applying an intervention can be used to construct machine learning models and then make the comparison. This comparison will then be judged against the quantitative objectives of the intervention. For this purpose, we distinguish two scenarios.

First Scenario: The Intervention Is Not Implemented in the Real System Yet: In this case, we only have access to historical data. A machine learning model can be trained on the historical data to predict the target variables in the future without the application of the intervention (i.e., assuming that no intervention is applied, what will happen in the future). It is also essential to understand what will happen if an intervention is implemented. For this purpose, various techniques can be used including: multi-agent simulation [41], physics-based modelling [42], [43], machine learning models [36], or a combination of these approaches. For example, in air quality modelling, physical meteorological models could be used that include Dispersion Models, Photochemical Models, or Receptor Models (See for example [42]). These mathematical models are based on the natural behaviour of physical quantities (concentration of the gas/particle, pressure, temperature, etc.), are time consuming to construct, and require iterative tuning of their parameters based on measured data.

The underlying principle of these prediction approaches is to raise assumptions on how the intervention would affect various features in the system and subsequently make predictions on the target variables. In this paper, we modify the historical dataset based on appropriately justified assumptions and train a second machine learning model for predicting target variables after the application of the intervention.

Second Scenario: The Intervention Is Already Implemented: In this case, data gathered and stored can be divided into two parts: historical data from the transportation system before the implementation of the intervention, and the most recent data from the system after the implementation of the intervention. A machine learning model can be trained on the historical data to predict the target variables in the future without the application of the intervention (i.e., assuming that no intervention was applied, what would have happened in the future). Then we compare the predicted target variables under

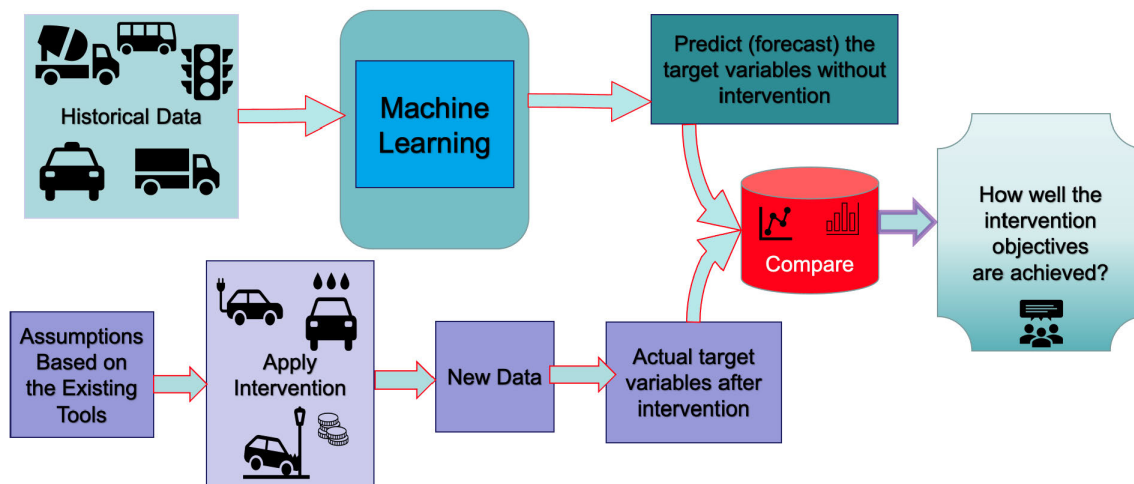


FIGURE 3. A framework for validating an intervention, and checking how well the objectives of the intervention are achieved.

no intervention with the target variables measured under the intervention. The new data obtained under intervention could also be used for improving the quality of the models built in the previous scenario, for instance by a better tuning of the model parameters.

C. PRE-PROCESSING DATASETS

We consider the datasets of the Urban observatory from the *air quality, weather, traffic* and *timestamp* themes. The variables for the air quality theme that we consider include CO, PM_{2.5}, PM₁, PM₁₀, PM₄, O₃, NO, and NO₂. The variables that we consider from the weather theme include Wind Direction, Wind Speed, Solar Radiation, Solar Diffuse Radiation, Pressure, Rain Duration, Rain ACC and Max Wind Speed. The variables from the traffic theme that we consider include Traffic Flow and Average Speed. We also have the timestamp theme that include Year, Month, Day, Hour, Minute, and Second, thus the measurements are taken every second.

Reducing NO₂ is the main objective of the intervention, thus we consider the availability of NO₂ data in each year. After analysing the NO₂ data gathered and stored by the Urban Observatory, we focus our work on the dataset for the year 2018 that has the largest number of measured values. The dataset of the year 2018 has over one million data entries, while the other years have a substantially smaller number of measurements. Therefore, we choose the year 2018 for training and validating the policy objective. Table 1 shows the statistics of the dataset for the year 2018. The table shows respectively the mean, standard deviation, minimum, 25th percentile (lower quartile), 50th percentile (Median), 75th percentile (upper quartile), and the maximum of each variable.

The pre-processing, computations, training, testing and visualisations of this paper are done using Python programming language. Our pre-processing takes into account measurements with obvious errors. For instance, any negative measurement of positive quantities is eliminated from the

dataset. Any measurement outside the bounds of quantities are also eliminated (e.g., any outlier sensor reading that is a few times higher than other readings). The same is applied to any data entry stored in the format of text that is supposed to be number.

The traffic flow measured and stored by Urban Observatory is the *total* number of buses, coaches, private cars, taxis, vans and heavy goods vehicles. The clean air zone affects these types of vehicles differently. For instance, it is designed to put restrictions and charge commercial vehicles without affecting private cars. In order to make accurate predictions on how the clean air zone reduces the NO₂ concentrations, it is essential to have separate datasets for traffic flow of different vehicle types. We divide the available dataset of the total traffic flow to four different traffic flow for different vehicle types including

- 1) buses and coaches,
- 2) heavy goods vehicles (HGVs),
- 3) cars and taxis, and
- 4) two-wheeled motor vehicles.

Since the available dataset includes only the total number and does not give separate numbers for four vehicle types, we use road traffic statistics published by the Department of Transport [44] to get the *percentages* of each vehicle type in Newcastle upon Tyne. According to this report, the percentages are as follows:

- Traffic flow of buses and coaches = 1.24% of the total traffic flow,
- Traffic flow of HGVs = 18.13% of the total traffic flow,
- Traffic flow of cars and taxis = 80% of the total traffic flow,
- Traffic flow of two-wheeled motor vehicles = 0.63% of the total traffic flow.

We visualise the available datasets in 2018 to extract useful knowledge and find suitable information. After this data exploration and visualisation, we filter the dataset to have the

TABLE 1. Statistics of the data for the year 2018. The table shows respectively the mean, standard deviation, minimum, 25th percentile (lower quartile), 50th percentile (Median), 75th percentile (upper quartile), and the maximum of each variable.

index	mean	std	min	25%	50%	75%	max
O3 (ppb)	269.53	332.38	0.09	23.32	137.87	438.24	2111.90
PM ₁ (μg/m ³)	3.65	4.01	0.18	1.20	2.15	4.43	44.06
PM _{2.5} (μg/m ³)	6.87	7.86	0.37	2.48	4.11	8.12	92.28
PM ₄ (μg/m ³)	9.99	8.46	0.18	4.78	7.59	12.00	134.13
PM ₁₀ (μg/m ³)	13.56	17.62	0.98	5.39	8.38	15.16	323.95
CO (ppm)	98.17	256.44	0.02	5.94	17.56	63.80	8091.09
NO (ppb)	58.99	38.44	7.29	31.07	49.88	77.10	247.87
NO ₂ (ppb)	55.22	49.77	0.34	19.32	41.02	74.37	313.73
Rain Duration (h)	9.48	6.41	0.00	4.00	9.00	14.00	50.76
Rain Acc.	3.99	2.99	0.0	0.90	2.87	5.10	15.54
Solar Radiation (w/m ²)	74.54	140.12	0.014	0.24	2.19	92.85	9.4.55
Wind Direction	196.17	36.31	106.33	169.55	207.02	225.28	276.40
Max Wind Speed (m/h)	3.34	1.91	0.00	1.92	2.93	4.402	16.05
Solar Diffuse Radiation (w/m ²)	62.83	91.19	1.01	1.48	11.19	94.08	531.15
Wind Speed (m/h)	9.63	8.69	1.73	6.26	7.99	10.42	152.87
Pressure (mb)	1005.81	13.64	969.07	993.39	1007.22	1017.30	1033.12
Traffic Flow (Passenger Car Units)	12.35	11.19	0.00	1.15	10.08	22.14	44.05
Average Speed (KmPH)	57.16	14.65	26.17	44.64	51.93	72.35	80.00

values of variables as time series. The dataset is also unified and transformed into an appropriate format to be compatible with machine learning methods.

From Continuous Values of NO₂ to Finite Number of Classes: The sensor measurements of NO₂ contain continuous quantities. We divide these values into five different ranges based on their risk for the general population [45]. The range of values is based on *parts per billion (ppb)*, which is defined as the number of units of mass of NO₂ per billion units of total mass. These ranges are

- 1) “Good” for NO₂ concentration between 0 – 50 ppb. The NO₂ concentrations in this range are expected to have no impact on health.
- 2) “Moderate” for NO₂ concentration between 51 – 100 ppb. This range of NO₂ is considered to be harmful for people who are sensitive to NO₂. These people should consider limiting extended outdoor exertion.
- 3) “Unhealthy for Sensitive Group” when the NO₂ concentration is between 101 – 150 ppb. This range of NO₂ is considered to be harmful for people with lung disease, for children and older people. They should *limit* extended outdoor exertion.
- 4) “Unhealthy” for NO₂ concentration between 151 – 200 ppb. Children, older people, and anyone with lung disease should *avoid* extended outdoor exertion. Anyone else should limit extended outdoor exertion.
- 5) “Very Unhealthy” for NO₂ concentration between 201 – 300 ppb. Children, older people, and anyone with

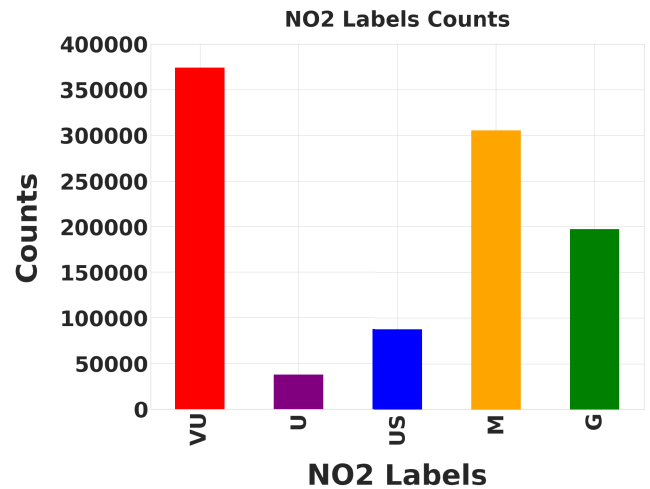


FIGURE 4. The number of NO₂ measurements in five different classes. The data has been ordered from most to least unhealthy. The classes are Very unhealthy (VU) with NO₂ concentration range between 201–300 ppb, Unhealthy (U) with NO₂ concentration range between 151–200 ppb, Unhealthy for sensitive group (US) with NO₂ concentration range between 101–150 ppb, Moderate (M) with NO₂ concentration range between 51–100 ppb, and Good (G) with NO₂ concentration range between 0–50 ppb.

lung disease should *avoid all* outdoor exertion. Anyone else should limit outdoor exertion.

Figure 4 represents the number of NO₂ measurements inside the above five classes. As it can be seen also from Figure 4, Very Unhealthy class has the highest number of measurements between these five classes. Therefore,

we expect that the machine learning models and the training learn the Very Unhealthy class quite well. On the other hand, the number of data entries for Unhealthy class is relatively smaller. We use averaging to change the time resolution of the measurements from second to hour. This will make the dataset a better representation of the hourly average quantities and make them more robust against measurement noises.

Since the clean air zone of Newcastle is not implemented yet, we only have access to historical data before the implementation of the intervention. To develop a machine learning model for predicting the concentration of NO₂ after the implementation of the clean air zone, we modify the historical dataset based on assumptions mentioned next.

- Implementation of the clean air zone of Newcastle will result in at least 20% reduction in the number of cars and taxis, 10% reduction in the number of buses and coaches, and 20% reduction in the number of HGVs.
- The implementation of the zone will result in an average reduction in the air pollution concentrations. To make this assumption quantitative and realistic, we use the Emissions Factors Toolkit (EFT) [37] published by the Department for Environment, Food & Rural Affairs (Defra) of the United Kingdom to estimate the average concentrations based on the traffic flow before and after the intervention. We take the difference between these two concentration estimations and deduct it from the original dataset. This means we modify the particles from the Air Quality Theme as follows: CO – 18, PM_{2.5} – 16, PM₁ – 10, PM₁₀ – 21, PM₄ – 23, O₃ – 29, NO – 18, NO_x – 24, and NO₂ – 25.

We have used these assumptions to demonstrate the applicability of our framework. These assumptions can be validated using approaches that study travel behaviour changes in response to the intervention [46], [47].

VI. MACHINE LEARNING MODELS

With the recent advances in artificial intelligence and big data, prediction methods based on machine learning models are becoming more and more common [48]. The main advantage of machine learning models is that their training do not require an understanding of air pollutants' physical or chemical properties. The structures and properties of machine learning models allow us to incorporate complex nonlinear relationships between the concentration of air pollutants and meteorological variables. In this section, we provide the details of the machine learning models used in our study. These models are Decision Trees [49], Light Gradient Boosting Machine [50], K-Nearest Neighbours [51], and Gradient Boosted Decision Tree [52]. We also give the test metrics that can be used to assess the performance of the learned model (accuracy, precision, recall, F1 score, and confusion matrix). We also discuss the details of one of the main contributions of the paper, which is time series analysis of the data for making predictions on the air quality.

The prediction problem can be formulated as a *classification* or *regression* problem depending on the range of quantities being predicted. These two are the main two categories of *supervised learning*, where the dataset contains labels that need to be predicted after training a model appropriately on the training dataset. In other words, it is clear what quantities need to be predicted and the values of these quantities is known for the train dataset. The difference between classification and regression is in the nature of predicted quantities: classification problem deals with discrete quantities or a finite number of classes (e.g., low/medium/high classification, or character recognition). An example of a classification problem is to predict mode of transport using an appropriate dataset. Mode of transport can take a finite number of different forms: bus, bike, car, taxi, and walking. Therefore, this is a classification problem. In contrast, the regression problem deals with continuous quantities that can take arbitrary values from an infinite set (e.g., price of a house, amount of rain, concentration of NO₂).

In the following subsections, we briefly discuss the underlying ideas of the classifiers used in this paper. We note that if the performance of the classifier is high (by appropriate selection of the hyper-parameters), this means that the classifier can capture the essential relations in the dataset and can provide more accurate predictions to be used by policy evaluators and policy makers. We emphasise that the general aim of our work is not to provide new machine learning algorithms but to show how current learning algorithms can be used in a framework for validating implementations of policy objectives. Our work is novel since previous research on validating objectives of transport policy interventions using data-driven methods is very limited as described in the related work section.

A. DECISION TREE

Decision Tree (DT) classifiers were initially developed in [53]. Since then, it has been used extensively as one of the most powerful classifiers. Recent use of DT classifiers in transport systems includes the work in [54] that predicts the mode choice behaviour of commuters and in [55] for automatic freeway incident detection. DT has a tree structure. At each node of the tree, the data is compared with a constant and depending on the direction of the comparison, one child node is selected. The leaf nodes of the tree hold the class labels. DT classifier assumes the labels to be a function of features. It tries to sequentially divide the space of features into two parts using comparison with a constant until the right label is identified. The depth of the tree is a hyper parameter that shows the required number of comparisons needed to assign a label to a data point.

B. LIGHT GRADIENT BOOSTING MACHINE

Light Gradient Boosting Machine (LGBM) classifier [50] is designed to be efficient and more effective for handling big data (large number of features and data instances). The

general idea of LGBM is to speed up learning and reduce the computational complexity by focusing on the training examples that result in a larger gradient. LGBM is used in transport applications, e.g., in [56] for predicting traffic crash severity.

C. K-NEAREST NEIGHBOURS

K-Nearest Neighbours (KNN) classifier was developed first in [57]. It assumes that data points that are near to each other in the feature space have the same label. The KNN algorithm predicts the label of a given data point as follows: it computes the distance of all the points from the current data point; it then sorts the computed distances from smallest to largest; it picks the first K smallest distances; it finally gets the labels of the selected K data points associated with those distances and returns the label with the highest repetition. The KNN classifier has been used extensively in transport applications including the work in [58] for short-term traffic forecasting and in [59] for imputation of missing traffic data.

D. GRADIENT BOOSTED DECISION TREES

Gradient Boosted Decision Tree (GBDT) combines multiple machine learning models (as weak learners) into a single machine learning model (as a strong learner) in an iterative fashion [60]. In GBDT, we use regression decision trees as weak models (a numerical value is assigned to each region of the feature space, which is the average of training data in that region). The loss function is also the log-loss function, which is then passed to a sigmoid function to find the predicted label. GBDT is used in transport applications for example in [61] to predict the occupancy of public transport vehicles.

E. JUSTIFICATION OF THE CHOICE OF CLASSIFIERS

The selection of DT, LGBM, KNN, and GBDT classifiers for identifying relevant data types associated with NO₂ can be justified as follows.

- DT classifiers are interpretable and non-parametric models that can capture complex relationships between features and the target variable [53]. They can help identify relevant data types by revealing which features are used to split the data and make predictions about NO₂ levels, highlighting the importance of these features.
- LGBM classifier is a gradient boosting framework that uses tree-based learning algorithms. It is an efficient and effective method for identifying relevant features, as it iteratively refines the model by focusing on the residuals of the previous iteration [50]. This results in a model that can emphasise the importance of certain features and their relationship with NO₂ levels.
- KNN classifier is a non-parametric, instance-based learning algorithm that can capture complex, nonlinear relationships between features and the target variable [57]. By analysing the proximity of data points in the feature space, KNN can identify relevant data types

that share similar characteristics, which may contribute to NO₂ levels.

- GBDT classifier is an ensemble learning method that combines multiple weak learners (usually decision trees) to build a strong predictive model [60]. By iteratively adding trees that correct the errors of the previous trees, GBDT can identify relevant data types by focusing on the features that contribute most to the prediction of NO₂ levels.

From the above four classifiers, GBDT and LGBM are developed recently and can have complex structures with larger number of hyper-parameters. The training of GBDT and LGBM could require more computational resources. DT and KNN are relatively simple and can be easily implemented. The KNN algorithm becomes slower for increasing data sizes (number of data points and features).

In summary, the combination of these methods provides a comprehensive approach to identifying relevant data types associated with NO₂ levels. These techniques cover a range of model types (linear, tree-based, instance-based) and feature selection methods (correlation, importance, proximity), allowing for a thorough exploration of the relationships between various data types and NO₂ levels.

F. TEST METRICS

For the initial assessment of the performance of the learning algorithms, we use accuracy, which is defined as $\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN})$, where TP, TN, FP, and FN indicates the number of predictions that are respectively true positives, true negatives, false positives, and false negatives. For Decision Tree classifier, we need to find the best maximum depth. This can be chosen by computing the accuracy of the classifier as a function of maximum depth. The best maximum depth can be selected such that the accuracy of the classifier is maximised.

There are also three other scores that can be used to assess the performance of the machine learning algorithms. Precision score is defined as $\text{Precision} = (\text{TP}) / (\text{TP} + \text{FP})$. Precision score gives the percent of correct predictions among all instances classified positive. Recall score is defined as $\text{Recall} = (\text{TP}) / (\text{TP} + \text{FN})$, which gives the percent of correct predictions among all instances that were actually positive. F1 score is defined as $\text{F1} = (2 \times \text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall})$. The F1 score can take its best value (one) when both precision and recall scores are equal to one. In worst case, it can be zero.

Confusion Matrix: The precision score defined above is very useful but does not contain all the information needed to assess the performance of a classification model. This is in particular important when the dataset is imbalanced (i.e., some of the labels may appear much less than other labels). The confusion matrix is a matrix that includes the statistics of the correct classes and predicted classes when the trained model is applied on the test dataset to make predictions.

A confusion matrix C is a square matrix with dimension n equal to the number of classes, where the entry C_{ij} of matrix

C shows the number of predictions predicted to be in group j and are known to be in group i . For binary classification (two classes), the numbers of true negatives is C_{00} , true positives is C_{11} , false positives is C_{01} and false negatives is C_{10} . The performance of a classification is judged to be good if the diagonal entries of the confusion matrix is close to 1 and the off-diagonal entries are close to 0.

G. TIME SERIES ANALYSIS AND PREDICTION

The dataset of air quality from the Urban Observatory includes measurements of the features as a function of time. These measurements can be seen as a sequence of data points that are sequential and are dependent along the time axis. Recurrent neural networks (RNNs) are a special class of neural networks with a particular structure designed for processing sequential data [48]. The basic idea of RNN is to introduce hidden states h_t that can encode some form of memory for capturing the essential information from previous data points in the sequence. The relation between hidden states h_t , features x_t , and labels y_t can be summarised with the equations

$$\begin{aligned} h_t &= f(x_t, h_{t-1}; b_h) \\ y_t &= g(h_t; b_y), \end{aligned}$$

where at each time point, the hidden state h_t is a function of current features x_t and previous hidden states h_{t-1} . The labels y_t are also functions of the hidden states h_t . The parameters b_h, b_y are learned from the data. The functions f, g are usually in the form of neurons that take the weighted sum of their inputs together with some appropriate activation functions.

Long Short-Term Memory (LSTM) networks are a type of RNNs for learning order dependence in sequence prediction problems [62]. An LSTM has four parts: a memory cell that keeps values over time periods of arbitrary length, and three gates (input, output, and forget gates) that regulate the information flow into and out of the cell. We have chosen LSTM networks for predicting NO₂ emissions instead of using Convolutional Neural Networks (CNN), Deep Neural Networks (DNN) or other types of networks due to the following reasons:

- Temporal dependencies: LSTM networks are specifically designed to handle time-series data and capture long-range dependencies, which are crucial when working with air quality data. Air quality variables like NO₂ emissions are influenced by past values and trends, making LSTMs more suitable for this task than DNNs, which do not explicitly model temporal dependencies.
- Handling of vanishing and exploding gradients: LSTMs are designed to overcome the vanishing and exploding gradient problems often encountered in training traditional RNNs. These issues make it difficult for RNNs to learn long-range dependencies in time-series data. LSTMs, with their gating mechanisms, can efficiently learn long-range dependencies without the gradient

problems, making them more appropriate for predicting NO₂ emissions.

- Robustness to missing data: In real-world air quality datasets, missing data is a common issue. LSTMs are more robust to missing data than DNNs due to their ability to maintain hidden states over time. This allows LSTMs to better handle gaps in the data and still provide accurate predictions.
- Better performance in practice: LSTMs have demonstrated superior performance in numerous time-series prediction tasks when compared to DNNs and other methods [62]. Their ability to model temporal dependencies and handle missing data makes them well-suited for predicting NO₂ emissions, which are influenced by a variety of factors that change over time.

Performance Metric: The accuracy of an LSTM model in making correct predictions on the dataset is evaluated by a metric called ‘‘Root Mean Square Error (RMSE),’’ defined as

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t|^2} \quad (1)$$

where y_t is the true output (vector of labels) and \hat{y}_t is the output predicted by the LSTM model.

VII. RESULTS AND DISCUSSION

After data analysis and pre-processing, suitable machine learning methods are selected for training and extracting the relevant datasets that are most important to the objective of the intervention. We divide our dataset randomly into 70% training data for learning the model and 30% test data for assessing the accuracy of the trained model.⁴ In our case study, we consider as reference point the objective of the intervention to be 10% reduction in the time duration when the concentration of NO₂ is unhealthy (NO₂ concentration above 100 ppb). In the next subsections, we discuss how to apply our framework presented in Section V to this clean air zone intervention.

A. APPLYING THE FIRST FRAMEWORK ON THE DATASET

We use the most popular machine learning models including DT, KNN, GBDT and LGBM to build models that can accurately capture the important information in the dataset, make accurate predictions, and help us to extract relevant important data types according to our first framework in Figure 2. A short description of these classification models can be found in Section VI. Note that the computation of Pearson Correlation for the framework can be done without the need for training a classification model, but the Feature Importance needs constructing a classification model first and then computing the importance values by shuffling the dataset.

⁴This is done using `train_test_split` function of `sklearn` package in Python.

TABLE 2. Normalised feature importance computed from different machine learning methods (LGBM, DT, KNN, GBDT), and the normalised correlation between NO₂ concentration and other measured variables. The variables belong to four themes, which is indicated with different colours in the first column (green, red, yellow, and dark blue). The variables in the air quality theme are in green colour, the variables in the timestamp theme are in red colour, the variables in the weather theme are in yellow colour, and the variables in the traffic theme are in dark blue colour. The cut-off value 1% is used to identify the most important features.

Measured Variables	LGBM	DT	KNN	GBDT	Correlation
O ₃	36.3	28.7	45.5	29.1	19.6
NO	1.4	1.8	4.3	2.1	3.1
CO	1.9	2.9	8.6	5.5	5.7
PM ₁	1	0.6	0.9	0.8	1.2
PM _{2.5}	1.9	1.8	0.9	1.6	1.6
PM ₄	1	0.6	0.9	0.8	0.5
PM ₁₀	2.9	1.2	0.9	1.6	2.1
Month	19.1	18.7	8.6	15	5.7
Day	4.8	13.5	6.9	8.7	9.4
Hour	1.2	2.6	1.9	2.2	2.4
Pressure	5.7	6.4	4.3	4.7	8.6
Wind Speed	1	1.8	0.9	0.8	0.8
Rain Duration	1	3.5	0.9	1.6	1.2
Max Wind Speed	1.9	1.2	1.4	1.2	4.9
Solar Diffuse Radiation	1	0.6	0.8	0.7	0.9
Wind Direction	2.9	1.2	4.3	0.8	10.6
Solar Radiation	1	1.8	2.6	1.6	1.2
Rain Acc.	1	0.2	0.2	0.1	0.9
Cars and taxis	13.4	11.7	10.3	23.6	16.4
Buses and coaches	1.4	2.1	1.7	2.21	3.1
HGVs	0.1	0.1	0.1	0.1	0.1
Two-wheeled motor vehicles	0.00	0.00	0.00	0.00	0.00
Average Speed	0.1	0.1	0.9	0.2	0.9

1) FIRST FRAMEWORK USING PEARSON CORRELATION

We compute the correlation between different features and the NO₂ concentration according to Equation (3) to find the most relevant variables related to the NO₂ according to our first framework in Figure 2. The result is presented in the right column of Table 2 after normalising the values. The feature importance does not have a specific range. To make them comparable, we have normalised the correlation coefficient and permutation feature importances and reported the percentages in the table. The numbers in the table are now in the range [0, 100]. These numbers show the relative importance of the features. We use the cut-off value 1% as a proof of concept to identify the most important features. In general, this cut-off value should be selected with respect to the size of the dataset and the available computational resources.

The measured variables are written in the left column of Table 2 and they are colour-coded to indicate each theme. The variables in the Air Quality Theme are in green colour, the variables in the Timestamp Theme are in red colour, the variables in the Weather Theme are in yellow colour, and the variables in the Traffic Theme are in dark blue colour.

The normalised correlation coefficient reported in the right column show that the two highest coefficients belong to the variables O₃ and the number of cars and taxis. We now look at coefficients across each theme separately. From the Air Quality Theme (green variables), O₃, NO, and CO have the highest correlation with NO₂ but PM₄ has a negligible correlation. In the Timestamp Theme (red variables),

Month and Day have the highest correlation. In the Weather Theme (yellow variables) Pressure and Wind Direction have the higher correlation, and in the Traffic Theme (dark blue variables), Cars and Taxis give the highest correlation. Note that although HGVs have a high emission factor, they have a negligible correlation because there is a small numbers of them.

2) FIRST FRAMEWORK USING FEATURE IMPORTANCE

After training the four classification models, we compute the Feature Importance values for all these models. The result is presented in Table 2. In each column, the relative importance values more than 1% are highlighted in magenta colour. For LGBM model, O₃ and Month have the highest importance. The following variables have the highest importance value in each theme: O₃ from the Air Quality Theme, Month from the Timestamp Theme, Pressure from the Weather Theme, and Car and taxis from the Traffic Theme.

For DT model, O₃ and Month have the highest importance. The following variables have the highest importance value in each theme: O₃ from the Air Quality Theme, Month from the Timestamp Theme, Pressure from the Weather Theme, and Car and Taxis from the Traffic Theme.

For KNN model, O₃ and Car and taxis have the highest importance. The following variables have the highest importance value in each theme: O₃ from the Air Quality Theme, Month from the Timestamp Theme, Pressure and Wind Direction from the Weather Theme, and Car and Taxis from the Traffic Theme.

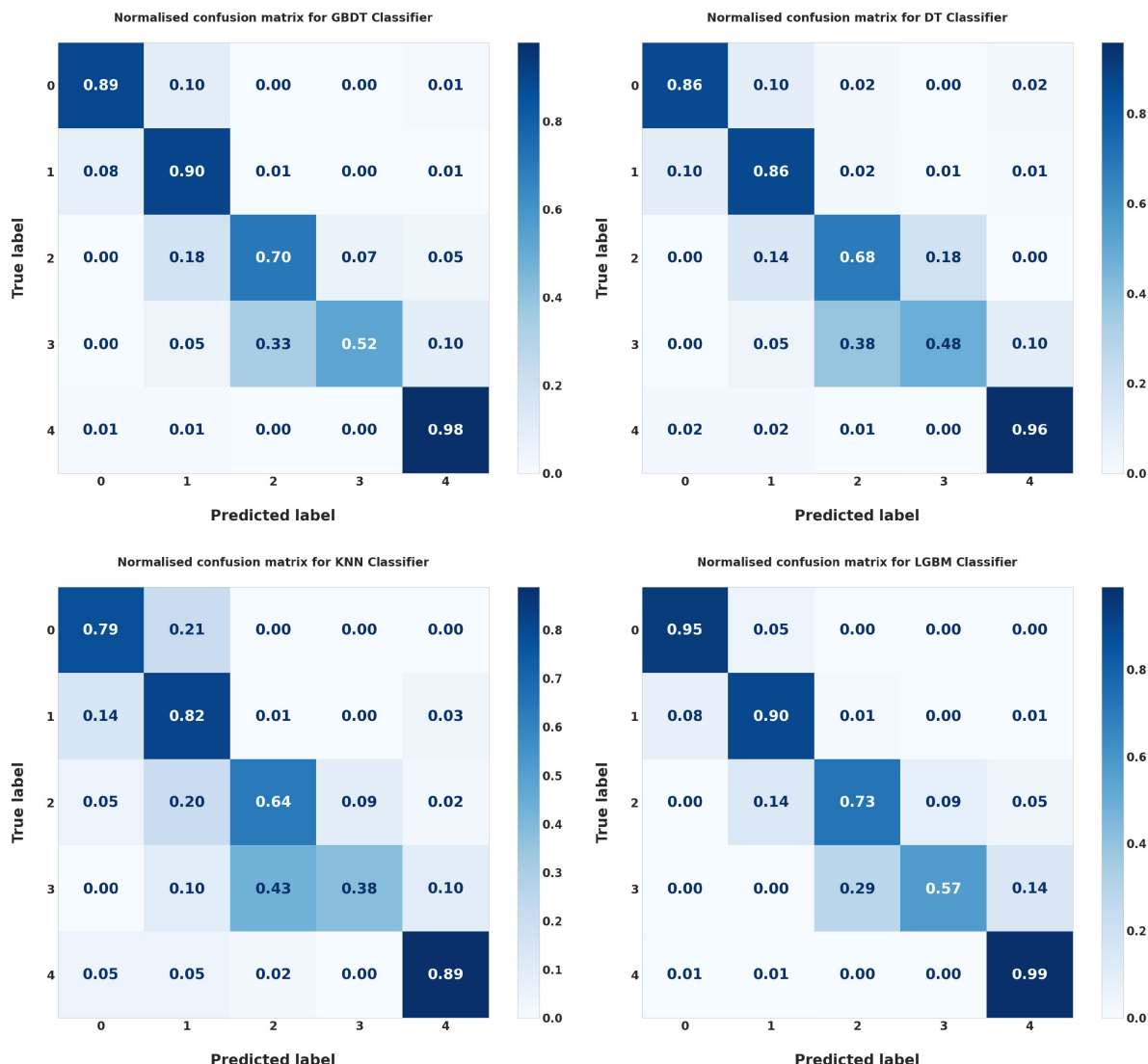


FIGURE 5. Normalised Confusion Matrix for the four classification models, from top left to bottom right for GBDT, DT, KNN, and LGBM models. The numbers associated with different classes are: 0 for Good, 1 for Moderate, 2 for Unhealthy Sensitive Group, 3 for Unhealthy, and 4 for Very Unhealthy.

For GBDT model, O₃ and Car and taxis have the highest importance. The following variables have the highest importance value in each theme: O₃ from the Air Quality Theme, Month from the Timestamp Theme, Pressure from the Weather Theme, and Car and Taxis from the Traffic Theme.

All these models share common conclusions about the importance of different variables in predicting NO₂ and validating or designing an intervention that involves reduction of NO₂. The conclusions of these models could be used in a voting mechanism to decide on the importance of features (similar to ensemble learning [63]). For this voting mechanism, we train machine learning models, compute the relative feature importances, and extract the features indicated as important by the majority of these models. For instance, most of these models used in our framework state that O₃, Month, Day, Pressure, and the Number of Cars and Taxis are

important. On the other hand, Wind Speed and Number of two-wheeled motor vehicles are less important in building a model for policy validation. These findings are also confirmed by the general intuitive observations about air quality: there is evidence of high correlation between NO₂ and O₃ [64], [65], [66]; the month is important as the air quality can get impacted due to the seasons and weather conditions; the day will impact the air quality as usually traffic volume might be higher during the working days and lower over the weekends. In our dataset, the average difference in traffic volume between weekends and the rest of the week is 35%.

There are also differences between the importance values computed using these models. This is mainly due to the nature of these algorithms that elements of randomness in their computations. The randomness is originating from the optimisation used to lean the model (exploring the parameter space of the model randomly), from the random initialisation

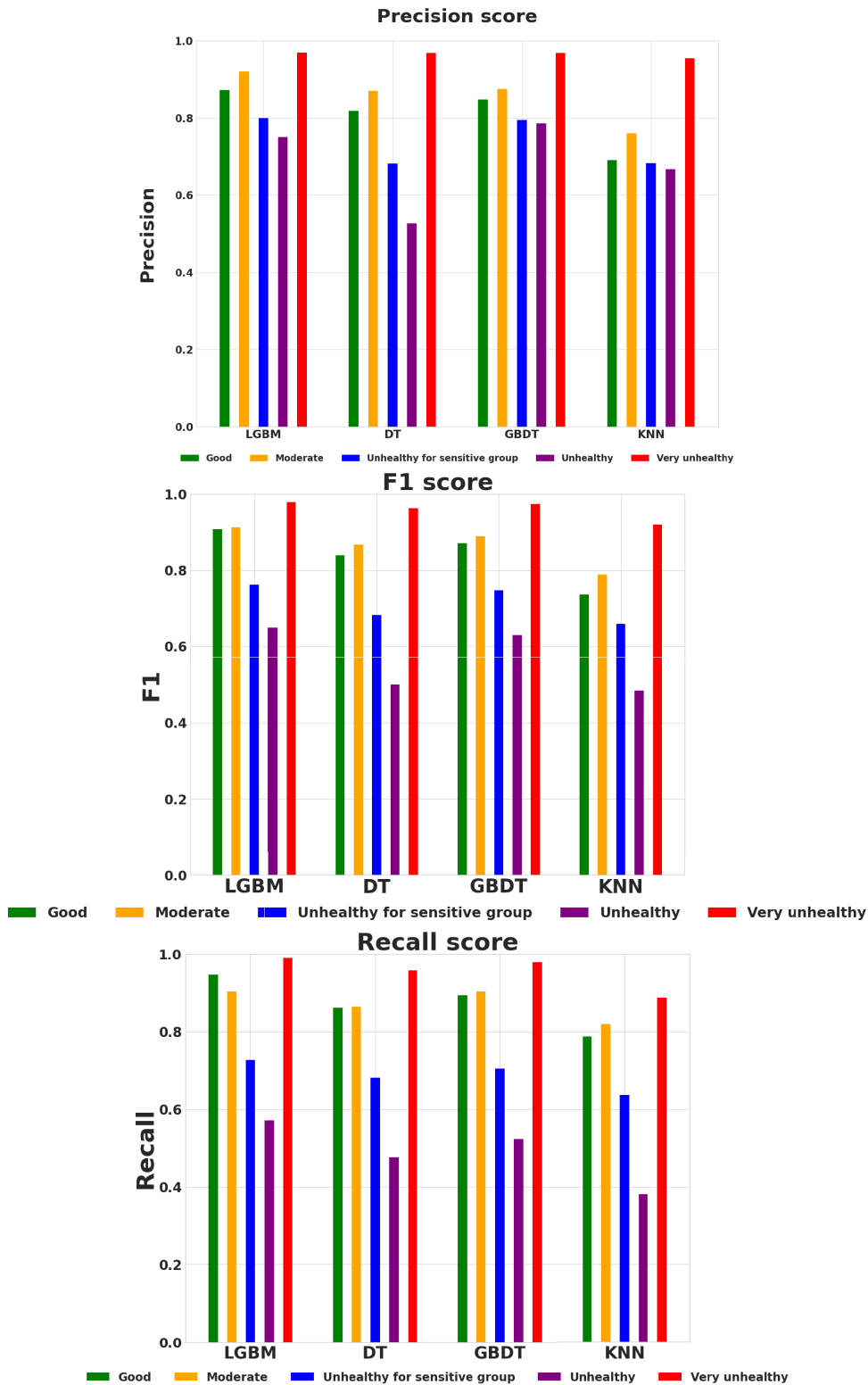


FIGURE 6. Precision, F1 and recall scores for each method.

of the optimisation, and from random splitting of the dataset to train-test data [67]. Therefore, any conclusion taken from the computations should be balanced with the accuracy of the constructed model. for instance, as we discuss in the rest of

this subsection, the accuracy of the LGBM model is lower compare to other models. Therefore, the results obtained based on this model should be discounted appropriately for making the final decision.

3) EVALUATION OF THE CLASSIFICATION MODELS

We have used Accuracy, Confusion Matrix, F1 score, Precision, and Recall to evaluate the obtained classification models (See Subsection VI-F for the definition of these metrics). The accuracy of the models are computed and repeated 10 times to account for different training and test split of the dataset. The average accuracy of the LGBM model is 88%, the average accuracy of DT model is 85%, the average accuracy of KNN model is 80%, and the average accuracy of GBDT model is 84%. The standard deviation of these accuracies is less than 1%. Among all these methods LGBM has the highest average accuracy and KNN model has the lowest average accuracy. This shows that the LGBM model can predict the correct class in almost 9 out of 10 cases. While this is an excellent outcome, the accuracy should be considered along other metrics to have the full understanding of accuracy in each class. Therefore, we also report the confusion matrix.

Figure 5 shows the Normalised Confusion Matrix for the four classification models, from top left to bottom right for GBDT, DT, KNN, and LGBM models. The numbers associated with different classes are: 0 for Good, 1 for Moderate, 2 for Unhealthy Sensitive Group, 3 for Unhealthy, and 4 for Very Unhealthy. Recall that for a good classification the diagonal entry of the confusion matrix should be close to one and the off-diagonal entries should be close to zero. As you can see from these Confusion Matrices, the models have learned the classes 1, 2, 3, 4 relatively well, but the performance is not good for class number 3. This is mainly due to the fact that our dataset has different number of data points in each class, and the number of data points in the Unhealthy class is much smaller than other classes. The accuracy of the obtained classification models can be improved by sampling techniques that deal with imbalanced datasets [68].

Figure 6 shows the comparison between the trained models using Precision, F1 score and Recall score. The figures provide these metrics for all five classes separately. The number of data points of classes are different as reported in Figure 4, the machine learning algorithms have different performances in capturing the relations in the data: the class Unhealthy with the smallest data points has the lowest score and the class Very Unhealthy with the largest data points has the highest score.

B. APPLYING THE SECOND FRAMEWORK FOR INTERVENTION EVALUATION

Since the clean air zone of Newcastle is not implemented yet, we use our framework in Figure 3 under the first scenario, where we only have access to historical data before the implementation of the intervention. We use LSTM model for training on the historical time series data and predict NO₂ concentrations (or its level) in the future without the application of the intervention. We also develop an LSTM model for predicting the concentration of NO₂ with the implementation of the clean air zone using the dataset modified according to Section V-C. The dataset of the year 2018 is divided into the first 10 months for training and evaluation,

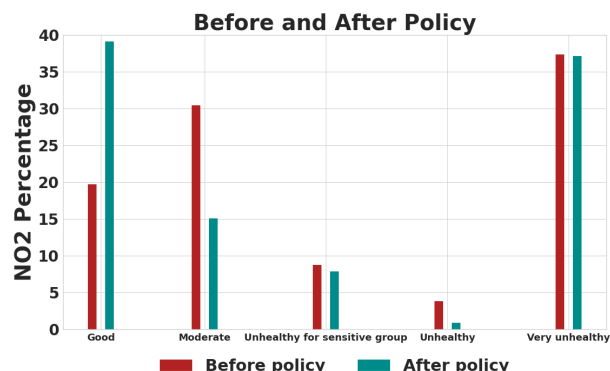


FIGURE 7. Percentages of each class of NO₂ (number of hourly NO₂ measurements in each class divided by the total number of hours). The results is for the year 2018 with and without the implementation of the clean air zone over the whole year. The number of NO₂ concentrations is reduced in all classes except Good class that is increased by 19.45%. The largest reduction is in the Moderate class (15.44% reduction) and the smallest reduction is in the Very Unhealthy class (0.18% reduction).

and the next two months for prediction. Figure 7 shows the percentages of each class of NO₂ predicted by the two LSTM model for the year 2018 with and without the intervention. According to Figure 7, the NO₂ concentrations of this year under no intervention will be 37.34% Very Unhealthy, 3.77% Unhealthy, 8.74% Unhealthy for Sensitive Group, 30.46% Moderate, and 19.69% Good. Figure 7 also shows that the NO₂ concentration of this year under intervention will be 37.16% Very Unhealthy, 0.84% Unhealthy, 7.84% Unhealthy for Sensitive Group, 15.02% Moderate, and 39.14% Good. As you can see, the number of NO₂ concentrations that are Very Unhealthy, Unhealthy, or Unhealthy for Sensitive Group and Moderate is reduced and the number of NO₂ concentrations that are Good is increased. The largest reduction is in the Moderate class (15.44% reduction) and the smallest reduction is in the Very Unhealthy class (0.18% reduction).

Figure 8 shows the LSTM prediction of NO₂ concentrations with and without the intervention in the available data points of months 11 and 12 (not used for training the models). The horizontal axes of these figures indicate the available time points ordered in a sequence. The vertical axes are the classes of NO₂ concentrations for these time points predicted by the LSTM models. Figure 9 shows the Cumulative Distribution Function (CDF) in the LSTM predictions with and without the intervention. According to the CDF, the classes 0 and 1 (good and moderate) have more amount than other three classes compared with the case of no intervention.

Outcome of Our Framework Applied to the Dataset: The number of NO₂ concentrations that are unhealthy is the sum of three classes: Very Unhealthy, Unhealthy, and Unhealthy for Sensitive Group. According to Figure 7, this is 37.34% + 3.77% + 8.74% = 49.85% without implementing the intervention and is 37.16% + 0.84% + 7.84% = 45.84% with the implementation of the intervention. This shows the total reduction of 49.85 - 45.84 = 4.01% and relative reduction of 4.01/49.85 = 8%. Therefore, the objective of the intervention which is 10% reduction will not be achieved

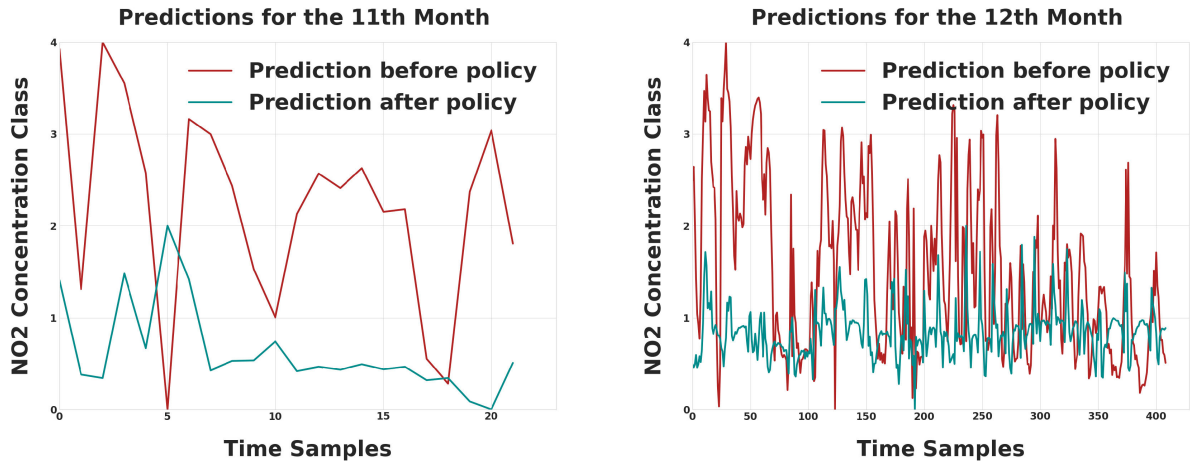


FIGURE 8. LSTM prediction for NO₂ concentrations for two months. The LSTM models are trained and evaluated on the available dataset from the first 10 months and are used for predicting the NO₂ concentration on the available data points of the next two months (more data points are available in the 12th month).

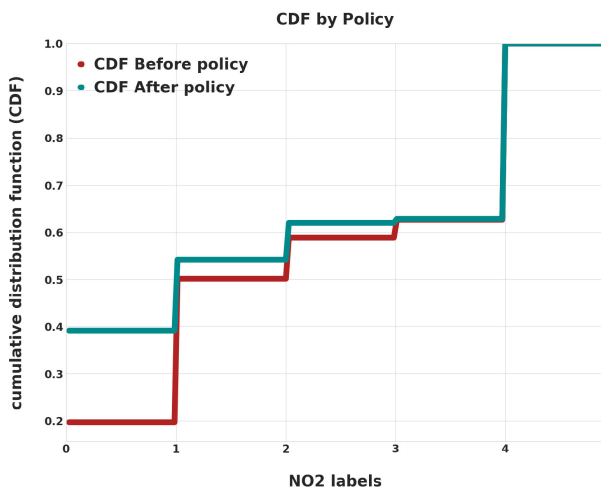


FIGURE 9. CDF of the LSTM method with and without the intervention.

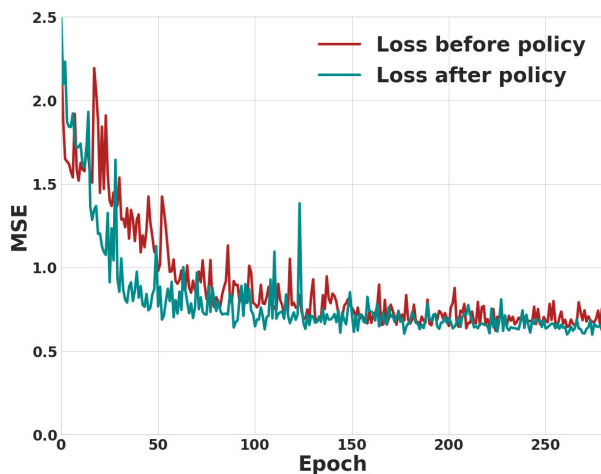


FIGURE 10. LSTM loss with and without the intervention as a function of epoch number (i.e., the number times that the learning algorithm will work through the entire training dataset for updating the model).

and additional intervention may be needed to reach the 10% reduction.

Time Series Model Evaluation: We use RMSE from Equation (1) to assess the performance of the learned LSTM models. The value of the RMSE for the model without the intervention is 0.946, while the RMSE for the model with the intervention is 0.850. This shows that the two model are performing similar to each other in terms of capturing the behaviour of data. Figure 10 shows the loss value of the training with and without the intervention. The loss starts from a high value and gradually decreases until converging to a fixed value while the learning algorithm finds the best parameters for the LSTM model.

VIII. CONCLUSION

In this paper, we discussed the use of machine learning methods for validating interventions in transportation systems using air quality and clean air zone intervention as a case study. We proposed a framework for finding data types that are relevant to the intervention objective. We also proposed a framework for validating the intervention, and checking how well the objectives of the intervention are achieved. We used the dataset from the Urban Observatory in Newcastle, United Kingdom, and considered an intervention related to clean air zone with the objective of reducing the concentration of NO₂. We developed an LSTM model for predicting the behaviour of the NO₂ without the implementation of the clean air zone.

In our first framework, we used the machine learning classifiers DT, KNN, GBDT and LGBM, and computed correlation coefficient and feature importances using these models. We then normalised these values to get the relative importance of the features. We used the cut-off value 1% as a proof of concept to identify the most important features. We showed that the constructed models share common conclusions about the importance of features in predicting NO₂, which could be used in a voting mechanism to decide on the importance of features. Our implementations also showed that among the selected learning models, LGBM performs best in capturing the relations in the dataset with accuracy 88%.

In our second framework, we used historical data of the year 2018 to model air quality in Newcastle upon Tyne both assuming no intervention is implemented and under the clean air zone implementation. The historical data from the first 10 months were used to build and evaluate the LSTM model, and the predictions are made for the last two months. The LSTM model can successfully predict the NO₂ concentration with root mean square error of 0.95. Our approach shows the use of machine learning methods in analysing and validating interventions in transportation systems. The role of machine learning can be summarised as predicting what is going to happen in the future if the policy is not implemented (using available historical data), and predicting the air quality and other related variables using transport behaviour changes in response to the implemented policy.

We used the term ‘framework’ since our approaches presented in the paper are high level and flexible, and can be applied to different policy objectives. The details and the choice of machine learning models can be decided depending on the specifics of the policy objective.

Our results are useful for the local authorities who are participating in the design and implementation of the clean air zone in Newcastle upon Tyne (e.g., Newcastle City Council). The clean air zone has come into effect with charges for non-compliant taxis, buses, coaches, and heavy good vehicles started from January 2023, and charges for vans and light goods vehicles will start from July 2023. It is important to assess the effectiveness of these charges in improving the air quality in Newcastle city centre and reducing the traffic-related pollution within the legal limits. Our frameworks help in modelling and understanding the relation between the gathered data, imposed charges, and reduction in air pollution at Newcastle. Our research contributes to a more sustainable urban environment by providing valuable insights into effective clean air zone interventions, which can improve air quality, reduce NO₂ concentrations, and promote sustainable transportation solutions. As a result, our work helps Newcastle city advance towards achieving its climate and air quality goals.

Our approach presented in this paper is currently focused on machine learning models that do not include any information from the pollutant’s physical/chemical models. It is also used as a proof of concept since the data after the implementation of the policy is not available yet. In the future, we plan to integrate our data-driven framework with physics-based models associated with an intervention to improve the performance and accuracy of our approach. We will also include optimisation methods in our framework to help designing better interventions for achieving the intervention objectives. We also note that the data quality is critical for drawing conclusions on the effectiveness of an intervention. There is additional work needed to improve the quality of the data stored in Urban Observatory, reduce the number of missing data points, and reduce the observation errors by calibrating sensors.

For future work, we plan to expand our data-driven framework to address additional sustainability-related challenges in urban transportation systems, such as promoting alternative modes of transportation and optimising the expansion of the electric vehicle charging infrastructure. This will allow us to design better interventions for achieving the intervention objectives and contribute to the sustainable ecosystem of the studied city. We plan to also analyse the data gathered from the clean air zone of Newcastle under the implementation of the zone and suggest improvements in implementing the zone (e.g., by adjusting the charges or categories of the cars).

APPENDIX

A. PYTHON PROGRAMMING LANGUAGE

We have selected *Python* programming language for the implementation of the approach of this paper that involves the process of large volumes of data. Python is open-source, and has an excellent documentation and a large community of developers. It also has a simple syntax that makes the coding faster when compared to other programming languages. Python comes with many libraries and frameworks. The most popular libraries are *NumPy* and *SciPy* for scientific computations, and *scikit* for data analysis. The framework *TensorFlow* is essential for machine learning projects. New libraries are being developed specifically for transportation applications (see e.g., the *TransBigData* package in [69]). We will use *Scikit-learn* library for training the machine learning models.

Pearson Correlation Coefficient: The correlation between different datasets and target variables can be computed using available data and statistical methods. Correlation coefficient gives a way to assess how much two variables are associated with each other. The correlation coefficient takes values in the interval $[-1, 1]$. A value of ± 1 indicates that two variables are dependent linearly. As the correlation coefficient goes towards 0, the relationship between the two variables will be weaker (they become independent of each other). A plus sign in the correlation coefficient shows positive relationship (increase in one variable will result in increase in another variable) and a minus sign shows a negative relationship (increase in one variable will result in decrease in another variable). More formally, for two random variables x, y , the correlation coefficient is defined as

$$\rho_{x,y} = \frac{\mathbb{E}[(x - m_x)(y - m_y)]}{\sigma_x \sigma_y}, \quad (2)$$

where \mathbb{E} is the expectation operator, m_x, m_y are the means and σ_x, σ_y are the variances of x, y . The correlation coefficient is computed using the following formula when a dataset $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ of size n is available for x, y :

$$\rho_{x,y} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}, \quad (3)$$

where \bar{x}, \bar{y} are the empirical means of x, y .

Permutation Feature Importance: The permutation feature importance was introduced in [70] for random forest models

and was extended in [71] to other machine learning models. In this metric, the importance of a feature is computed by first training the model on the train dataset, then permuting the feature and computing the increase in the prediction error of the model. If a feature is important in making predictions, the prediction error should increase after permutation. If a feature is unimportant, the change in the prediction error will be by doing the permutation.

REFERENCES

- [1] *Clean Air Strategy*, DEFRA, London, U.K., 2019.
- [2] T.-M. Chen, W. G. Kuschner, J. Gokhale, and S. Shofer, "Outdoor air pollution: Nitrogen dioxide, sulfur dioxide, and carbon monoxide health effects," *Amer. J. Med. Sci.*, vol. 333, no. 4, pp. 249–256, Apr. 2007.
- [3] C. Holman, R. Harrison, and X. Querol, "Review of the efficacy of low emission zones to improve urban air quality in European cities," *Atmos. Environ.*, vol. 111, pp. 161–169, Jun. 2015.
- [4] J. Zhang, F.-Y. Wang, K. Wang, W.-H. Lin, X. Xu, and C. Chen, "Data-driven intelligent transportation systems: A survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 4, pp. 1624–1639, Dec. 2011.
- [5] A. I. Torre-Bastida, J. Del Ser, I. Laña, M. Ilardia, M. N. Bilbao, and S. Campos-Cordobés, "Big data for transportation and mobility: Recent advances, trends and challenges," *IET Intell. Transp. Syst.*, vol. 12, no. 8, pp. 742–755, Oct. 2018.
- [6] A. Urbanek, "Data-driven transport policy in cities: A literature review and implications for future developments," in *Scientific and Technical Conference Transport Systems Theory and Practice*. Cham, Switzerland: Springer, 2018, pp. 61–74.
- [7] T. Teoh, B. van Berne, I. Hindriks, R. Waldenfels, T. Eichhorn, T. Ivanov, M. Hong, R. Akerkar, M. Benkic, and J. Sangwan, "Leveraging big data for managing transport operations," LEMO Project, Horizon Res. Innov. Programme, Eur. Union, Tech. Rep. D3.2, 2019.
- [8] K. M. Hoen, T. Tan, J. C. Fransoo, and G.-J. van Houtum, "Switching transport modes to meet voluntary carbon emission targets," *Transp. Sci.*, vol. 48, no. 4, pp. 592–608, 2014.
- [9] D. Sarabia-Jacome, C. E. Palau, M. Esteve, and F. Boronat, "Seaport data space for improving logistic maritime operations," *IEEE Access*, vol. 8, pp. 4372–4382, 2020.
- [10] G. Dai, C. Ma, and X. Xu, "Short-term traffic flow prediction method for urban road sections based on space–time analysis and GRU," *IEEE Access*, vol. 7, pp. 143025–143035, 2019.
- [11] Y. Ma, Z. Zhang, and A. Ihler, "Multi-lane short-term traffic forecasting with convolutional LSTM network," *IEEE Access*, vol. 8, pp. 34629–34643, 2020.
- [12] H.-Y. Li, W. Xu, Y. Cui, Z. Wang, M. Xiao, and Z.-X. Sun, "Preventive maintenance decision model of urban transportation system equipment based on multi-control units," *IEEE Access*, vol. 8, pp. 15851–15869, 2020.
- [13] Y. Zhao, J. Zhou, Y. Fan, and H. Kuang, "An expected utility-based optimization of slow steaming in sulphur emission control areas by applying big data analytics," *IEEE Access*, vol. 8, pp. 3646–3655, 2020.
- [14] X. Lu, K. Ota, M. Dong, C. Yu, and H. Jin, "Predicting transportation carbon emission with urban big data," *IEEE Trans. Sustain. Comput.*, vol. 2, no. 4, pp. 333–344, Oct. 2017.
- [15] N. Sharma and P. K. De, "Application of machine learning to predict CO₂ emission from transport sector to mitigate climate change," in *Towards Net-Zero Targets: Usage of Data Science for Long-Term Sustainability Pathways*. Singapore: Springer, 2022, pp. 197–218.
- [16] *The Largest Set of Publicly Available Real Time Urban Data in the U.K.*, Urban Observatory, New Delhi, India, 2022. Accessed: Feb. 20, 2022.
- [17] F. Farhadi, R. Palacin, and P. Blythe, "Data-driven framework for validating policies: Air quality case study," in *Proc. IEEE 25th Int. Conf. Intell. Transp. Syst. (ITSC)*, Oct. 2022, pp. 1–6.
- [18] P. James, R. Das, A. Jalosinska, and L. Smith, "Smart cities and a data-driven response to COVID-19," *Dialogues Hum. Geogr.*, vol. 10, no. 2, pp. 255–259, Jul. 2020.
- [19] *Charging Clean Air Zone Class C in Newcastle*, Newcastle City Council, Tyne, U.K., 2022. Accessed: Feb. 20, 2022.
- [20] *Clean Air Zone Framework: Principles for Setting Up Clean Air Zones in England*, Defra Joint Air Qual. Unit, DEFRA and DFT, London, U.K., Feb. 2020.
- [21] *Greater Manchester Clean Air Zone*, Manchester City Council, Manchester, U.K., 2022. Accessed: Feb. 20, 2022.
- [22] *Air Quality in Europe—2020 Report*, Eur. Environ. Agency, Luxembourg, U.K., 2020.
- [23] R. B. Ellison, S. P. Greaves, and D. A. Hensher, "Five years of London's low emission zone: Effects on vehicle fleet composition and air quality," *Transp. Res. D, Transp. Environ.*, vol. 23, pp. 25–33, Aug. 2013.
- [24] *London Low Emission Zone: Impacts Monitoring Baseline Report*. London, Transp. London, London, U.K., 2008.
- [25] F. Karaca, "Determination of air quality zones in Turkey," *J. Air Waste Manage. Assoc.*, vol. 62, no. 4, pp. 408–419, Apr. 2012.
- [26] A. Sturman and P. Zawar-Reza, "Application of back-trajectory techniques to the delimitation of urban clean air zones," *Atmos. Environ.*, vol. 36, no. 20, pp. 3339–3350, Jul. 2002.
- [27] S. Yearley, S. Cinderby, J. Forrester, P. Bailey, and P. Rosen, "Participatory modelling and the local governance of the politics of U.K. Air pollution: A three-city case study," *Environ. Values*, vol. 12, no. 2, pp. 247–262, May 2003.
- [28] H. Yeh, "The effects of successful ICT-based smart city services: From citizens' perspectives," *Government Inf. Quart.*, vol. 34, no. 3, pp. 556–565, Sep. 2017.
- [29] D. Iskandaryan, F. Ramos, and S. Trilles, "Air quality prediction in smart cities using machine learning technologies based on sensor data: A review," *Appl. Sci.*, vol. 10, no. 7, p. 2401, Apr. 2020.
- [30] L.-J. Chen, Y.-H. Ho, H.-C. Lee, H.-C. Wu, H.-M. Liu, H.-H. Hsieh, Y.-T. Huang, and S.-C. C. Lung, "An open framework for participatory PM_{2.5} monitoring in smart cities," *IEEE Access*, vol. 5, pp. 14441–14454, 2017.
- [31] M. Catalano and F. Galatioto, "Enhanced transport-related air pollution prediction through a novel metamodel approach," *Transp. Res. D, Transp. Environ.*, vol. 55, pp. 262–276, Aug. 2017.
- [32] R. Navares and J. L. Aznarte, "Predicting air quality with deep learning LSTM: Towards comprehensive models," *Ecol. Informat.*, vol. 55, Jan. 2020, Art. no. 101019.
- [33] Y. Rybarczyk and R. Zalakeviciute, "Assessing the COVID-19 impact on air quality: A machine learning approach," *Geophys. Res. Lett.*, vol. 48, no. 4, pp. 1–11, Feb. 2021.
- [34] A. J. Turner, J. Kim, H. Fitzmaurice, C. Newman, K. Worthington, K. Chan, P. J. Wooldridge, P. Köehler, C. Frankenberg, and R. C. Cohen, "Observed impacts of COVID-19 on urban CO₂ emissions," *Geophys. Res. Lett.*, vol. 47, no. 22, pp. 1–6, 2020.
- [35] M. Castelli, F. M. Clemente, A. Popovič, S. Silva, and L. Vanneschi, "A machine learning approach to predict air quality in California," *Complexity*, vol. 2020, pp. 1–23, Aug. 2020.
- [36] A. Suleiman, M. R. Tight, and A. D. Quinn, "Applying machine learning methods in managing urban concentrations of traffic-related particulate matter (PM₁₀ and PM_{2.5})," *Atmos. Pollut. Res.*, vol. 10, no. 1, pp. 134–144, Jan. 2019.
- [37] *Emissions Factors Toolkit*, Defra, London, U.K., 2022. Accessed: Feb. 20, 2022.
- [38] A. Orun, D. Elizondo, E. Goodyer, and D. Paluszczyszyn, "Use of Bayesian inference method to model vehicular air pollution in local urban areas," *Transp. Res. D, Transp. Environ.*, vol. 63, pp. 236–243, Aug. 2018.
- [39] Doreswamy, K. S. Harishkumar, K. M. Yogesh, and I. Gad, "Forecasting air pollution particulate matter (PM_{2.5}) using machine learning regression models," *Proc. Comput. Sci.*, vol. 171, pp. 2057–2066, Jan. 2020.
- [40] S. Vosough, H. Poorzahedy, and R. Lindsey, "Predictive cordon pricing to reduce air pollution," *Transp. Res. D, Transp. Environ.*, vol. 88, Nov. 2020, Art. no. 102564.
- [41] A. Doniec, R. Mandiau, S. Piechowiak, and S. Espié, "A behavioral multi-agent model for road traffic simulation," *Eng. Appl. Artif. Intell.*, vol. 21, no. 8, pp. 1443–1454, 2008.
- [42] R. A. Pielke and M. Uliasz, "Use of meteorological models as input to regional and mesoscale air quality models—Limitations and strengths," *Atmos. Environ.*, vol. 32, no. 8, pp. 1455–1466, Apr. 1998.
- [43] H. Rastgoftar and J.-B. Jeannin, "A physics-based finite-state abstraction for traffic congestion control," 2021, *arXiv:2101.07865*.
- [44] *Road Traffic Statistics*, Dept. Transp., London, U.K., 2022. Accessed: Feb. 20, 2022.
- [45] *Air Quality Guide for Nitrogen Dioxide*, Office Air Radiat. (6301A), USA, vol. 1, no. 2, Feb. 2011.

- [46] B. Conlan, A. Fraser, M. Vedrenne, J. Tate, and A. Whittles, "Evidence review on effectiveness of transport measures in reducing nitrogen dioxide," DEFRA, Didcot, U.K., Tech. Rep. AQ0959, May 2016.
- [47] B. Arnott, L. Rehackova, L. Errington, F. F. Sniehotta, J. Roberts, and V. Araujo-Soares, "Efficacy of behavioural interventions for transport behaviour change: Systematic review, meta-analysis and intervention coding," *Int. J. Behav. Nutrition Phys. Activity*, vol. 11, no. 1, pp. 1–23, Dec. 2014.
- [48] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep Learning*, vol. 1, no. 2. Cambridge, MA, USA: MIT Press, 2016.
- [49] L. Breiman, *Classification and Regression Trees*. Evanston, IL, USA: Routledge, 2017.
- [50] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, "LightGBM: A highly efficient gradient boosting decision tree," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 3149–3157.
- [51] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Trans. Inf. Theory*, vol. IT-13, no. 1, pp. 21–27, Jan. 1967.
- [52] L. Mason, J. Baxter, P. Bartlett, and M. Frean, "Boosting algorithms as gradient descent in function space," in *Proc. NIPS*, vol. 12, 1999, pp. 512–518.
- [53] R. Messenger and L. Mandell, "A modal search technique for predictive nominal scale multivariate analysis," *J. Amer. Stat. Assoc.*, vol. 67, no. 340, pp. 768–772, Dec. 1972.
- [54] C. R. Sekhar, Minal, and E. Madhu, "Mode choice analysis using random Forrest decision trees," *Transp. Res. Proc.*, vol. 17, pp. 644–652, Jan. 2016.
- [55] S. Chen and W. Wang, "Decision tree learning for freeway automatic incident detection," *Expert Syst. Appl.*, vol. 36, no. 2, pp. 4101–4105, Mar. 2009.
- [56] J. Niyogisubizo, L. Liao, F. Zou, G. Han, E. Nziyumva, B. Li, and Y. Lin, "Predicting traffic crash severity using hybrid of balanced bagging classification and light gradient boosting machine," *Intell. Data Anal.*, vol. 27, no. 1, pp. 79–101, Jan. 2023.
- [57] E. Fix and J. L. Hodges, "Discriminatory analysis. Nonparametric discrimination: Consistency properties," *Int. Stat. Rev./Revue Internationale de Statistique*, vol. 57, no. 3, pp. 238–247, 1989.
- [58] F. Su, H. Dong, L. Jia, Y. Qin, and Z. Tian, "Short-term traffic forecasting using self-adjusting k-nearest neighbours," *IET Intell. Transp. Syst.*, vol. 12, no. 1, pp. 41–48, Feb. 2018.
- [59] Y. Wang, Y. Xiao, J. Lai, and Y. Chen, "An adaptive k nearest neighbour method for imputation of missing traffic data based on two similarity metrics," *Arch. Transp.*, vol. 54, no. 2, pp. 59–73, 2020.
- [60] J. Elith, J. R. Leathwick, and T. Hastie, "A working guide to boosted regression trees," *J. Animal Ecol.*, vol. 77, no. 4, pp. 802–813, Jul. 2008.
- [61] F. Gallo, F. Corman, and N. Sacco, "Real-time occupancy predictions of public transport vehicles," in *Proc. 22nd Swiss Transp. Res. Conf. (STRC)*, 2022, pp. 1–19.
- [62] F. A. Gers, N. N. Schraudolph, and J. Schmidhuber, "Learning precise timing with LSTM recurrent networks," *J. Mach. Learn. Res.*, vol. 3, no. 1, pp. 115–143, 2003.
- [63] R. Polikar, "Ensemble learning," in *Ensemble Machine Learning*. New York, NY, USA: Springer, 2012, pp. 1–34.
- [64] S. Han, H. Bian, Y. Feng, A. Liu, X. Li, F. Zeng, and X. Zhang, "Analysis of the relationship between O₃, NO and NO₂ in Tianjin, China," *Aerosol Air Quality Res.*, vol. 11, no. 2, pp. 128–139, 2011.
- [65] L. Dimitrievici, D.-E. Constantin, and L. Moraru, "The analysis of the correlations between NO₂ column, O₃ column and UV radiation at global level using space observations," *AIP Conf.*, vol. 1796, no. 1, 2017, Art. no. 030009.
- [66] Y. Zhao, K. Zhang, X. Xu, H. Shen, X. Zhu, Y. Zhang, Y. Hu, and G. Shen, "Substantial changes in nitrogen dioxide and ozone after excluding meteorological impacts during the COVID-19 outbreak in mainland China," *Environ. Sci. Technol. Lett.*, vol. 7, no. 6, pp. 402–408, Jun. 2020.
- [67] I. H. Witten and E. Frank, "Data mining: Practical machine learning tools and techniques with Java implementations," *ACM SIGMOD Rec.*, vol. 31, no. 1, pp. 76–77, 2002.
- [68] M. Kuhn and K. Johnson, *Applied Predictive Modeling*, vol. 26. New York, NY, USA: Springer, 2013.
- [69] Q. Yu and J. Yuan, "TransBigData: A Python package for transportation spatio-temporal big data processing, analysis and visualization," *J. Open Source Softw.*, vol. 7, no. 71, p. 4021, Mar. 2022.
- [70] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
- [71] A. Fisher, C. Rudin, and F. Dominici, "All models are wrong, but many are useful: Learning a variable's importance by studying an entire class of prediction models simultaneously," *J. Mach. Learn. Res.*, vol. 20, no. 177, pp. 1–81, 2019.



FARZANEH FARHADI received the B.Sc. degree (Hons.) in software engineering in The Netherlands and the M.Sc. degree (Hons.) in advanced computer science in U.K. She is currently pursuing the Ph.D. degree with the School of Engineering, Newcastle University, U.K.

She has over eight years of experience working with international companies in software industry (Sage, Scott Logic, Bol, Mendix, and OCLC). Her Ph.D. study in the field of intelligent transport systems, supported by the U.K. EPSRC Research Council in collaboration with Arup Group Ltd. Her M.Sc. thesis was on developing deep learning models for classification and segmentation of health-related data. Her research interests include develop data-driven solutions in transport systems using machine learning and optimization methods with case studies in electric vehicle charging infrastructure and clean air zone.



ROBERTO PALACIN received the M.Sc. degree in rail systems engineering from the University of Sheffield and the Ph.D. degree in energy efficiency of urban rail systems from Newcastle University.

He has been the Director of Mechanical Engineering and Marine Technology, since 2020. He is currently a Professor in transport futures with Newcastle University, U.K. He has contributed to research projects on subjects, such as development of innovative railway concepts, urban mobility, energy optimization of rail systems, intermodality of the European rail networks, and development of modular concepts for high-speed rails. His research interests include systems-thinking applied to mobility, vehicle interiors, the development of ergonomic and design-led transport environments, and improving railways energy efficiency.



PHIL BLYTHE is currently a Professor in intelligent transport systems (ITS) with Newcastle University, U.K. He is also a Chartered Engineer, the Vice President, and a fellow of the Institution of Engineering and Technology (IET). He is also the chairs of the IET's Transport Policy Panel. He was the Director of the Transport Operations Research Group (2002–2015) and a former Chief Scientific Adviser with the Department of Transport (2015–2021). He is internationally recognized as

an early pioneer of ITS research. His academic focus has been the development of ITS—the use of information, communications, and computing technology applied to transport. His research interests include road to vehicle communications, road user charging systems, ITS for assistive mobility, smartcards and radio frequency identification, electro-mobility, and future intelligent infrastructure. His research bridges the technology-policy gap in terms of what technologies may evolve to meet future policy objectives or influence future policy thinking to meet the challenges.

He was an elected fellow of the Royal Academy of Engineering, in 2020. He was awarded the Reece-Hills Medal for a lifetime personal contribution to ITS, in 2012, and was awarded the honour of the Order of Commander of the British Empire, in 2022, for services to Science and Engineering in Transport and Government.

...