This article has been accepted for publication in IEEE Open Journal of Vehicular Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/OJVT.2024.3484956

; revised .

Digital Twin-empowered Green Mobility Management in Next-Gen Transportation Networks

Kubra Duran[1,2], Graduate Student Member, IEEE, Lal Verda Cakir[1,2], Graduate Student Member, IEEE, Achille Fonzone[3], Trung Q. Duong[4,5], Fellow, IEEE, and Berk Canberk[1], Senior Member, IEEE

[1]School of Computing, Engineering and Built Environment, Edinburgh Napier University, Edinburgh EH10 5DT UK
[2]BTS Group, Istanbul 34467 Turkey
[3]Transport Research Institute, Edinburgh Napier University, EH10 5DT Edinburgh UK
[4]Faculty of Engineering and Applied Science, Memorial University of Newfoundland, St. John's, NL A1C 5S7 Canada
[5]School of Electronics, Electrical Engineering and Computer Science, Queen's University Belfast, BT7 1NN Belfast UK

CORRESPONDING AUTHOR: Kubra Duran (e-mail: kubra.duran@napier.ac.uk).

ABSTRACT Evolving transportation networks need seamless integration and effective infrastructure utilisation to form the next-generation transportation networks. Also, they should be capable of capturing the traffic flow data at the right time and promptly applying sustainable actions toward emission reduction. However, traditional transportation networks cannot handle right-time updates and act upon the requirements in dynamic conditions. Here, Digital Twin (DT) enables the development of enhanced transportation management via robust modelling and intelligence capabilities. Therefore, we propose a DT-empowered Eco-Regulation (DTER) framework with a novel twinning approach. We define a transport-specific twin sampling rate to catch right-time data in a transportation network. Besides, we perform emission prediction using Multi-Layer Perceptron (MLP), Bidirectional Long Short-Term Memory (Bi-LSTM), and BANE embeddings. We perform Laplacian matrix analysis to cluster the risk zones regarding the emissions. Thereafter, we recommend actions by setting the number of vehicle limits of junctions for high-emission areas according to the outputs of Q-learning. In summary, DTER takes control of the emission with its transport-specific twin sampling rate and automated management of transportation actions by considering the emission predictions. We note DTER achieves 19% more successful right-time data capturing, with 30% reduced query time. Moreover, our hybrid implementation of intelligent algorithms for emission prediction resulted in higher accuracy when compared to baselines. Lastly, the autonomous recommendations of DTER achieved $\sim$ 20% decrease in emissions by presenting an effective carbon tracing framework.

INDEX TERMS autonomous traffic management, digital twin, reinforcement learning, twin sampling rate.

I. INTRODUCTION

In recent years, various transportation issues are frequently encountered stemming from the mobility of vehicles and pedestrians. Besides, additional problems, such as congestion, accidents, and inefficient service delivery, appear due to the lack of proper planning and management [1]. These problems become complicated due to growing traffic volume, thus leading to increasing operational complexity and reduced efficiency in traffic management [2]. Furthermore, there is a pressing need to reduce road transportation emissions. The fact that 27% of UK-wide emissions are caused by transportation urges us to make immediate changes to be applied [3]. However, the current transport management systems prioritize improving traffic flow in the decision-making process rather than reducing emission levels. Although route management schemes [4]–[7] have been widely researched, they require significant changes in the infrastructure of road networks, vehicles, and existing communication infrastructure. Therefore, adopting and integrating new technologies is the need of the hour to present sustainable next-generation transportation systems by hitting the emission reduction goal. Regarding that, the efforts and implementation of Digital Twins (DTs) in intelligent transportation have been started [8] especially for traffic safety and mobility [9]–[11], and network traffic prediction [12], [13]. In this context, the DTs present virtual

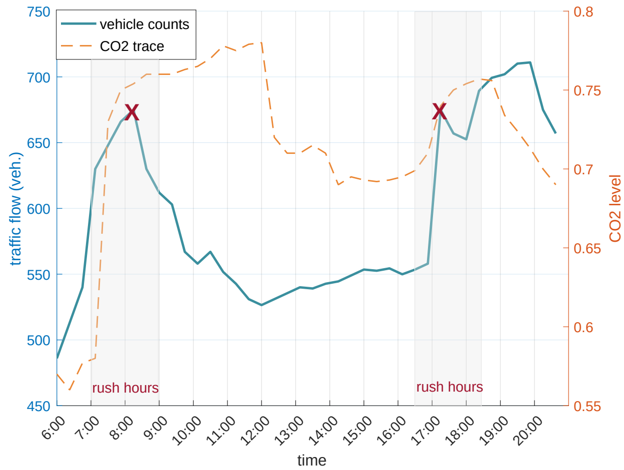


FIGURE 1. Transport network traffic flow and emission trace characteristics.

mirrors of the physical transportation network to perform what-if analyses.

A. MAIN CHALLENGES IN TRANSPORTATION NETWORK MANAGEMENT

- *Right-time data capturing to ensure efficient management:* Right-time data balances the benefits of real-time data and resource usage (processing time for data querying, CPU usage, etc.) of a transport management system when utilized in the DT context. As seen in Fig. 1, the traffic flow, in other words, the number of vehicles in a transportation network, is dynamic and depends on the rush hours of the day. Within these time slots, the traffic flow shows a sharp increase. These changes, especially the peak points, result in inaccurate and false representations of the physical transport environment due to the inability to catch the data at the right times, leading to invalid predictions and unfit traffic regulations.
- *Inadequacy of utilizing temporal and spatial features in $CO2$ prediction:* As seen in Fig. 1, the trace of the $CO2$ levels, shown by orange lines, increases with the pattern of dynamical changes in traffic flow. At this point, predicting emission levels in a transport network is advantageous for efficient management and keeping the emissions below the desired level (shown in the purple dashed line in the figure). However, considering only temporal features is insufficient for accurate emissions prediction; location information is also necessary.
- *Lack of intelligent recommendations to control emission levels via autonomous-driven actions:* Updating traffic actions to control transportation networks is an effective remedy for sustainability. In this regard, traditional transportation management is mostly driven by rule-based systems. However, this is not sufficient to draw conclusions about the transport network's future $CO2$ levels and recommend potential actions.

B. MOTIVATION

As an actual blueprint of the physical environment with a one-system-fits-all perspective, DTs become a key enabler to meet the design requirements of sustainable next-generation transportation networks. Thanks to its robust modelling and cognitive capabilities, the transportation network is replicated with real-time data and the risky zones can be identified in a robust manner. Based on these, our motivation for this work is three-fold:

Transport-specific twinning for right-time data capturing: To hit the dynamic changes in DT-based transport management at the right times while preserving the system resources, the definition of use-case-specific data capturing is required. This will highly improve the efficiency of transport services and resource utilization. For instance, with a right-time DT system, the transport management centre can use the latest dynamic changes in information and be fully aware of the information's freshness [14] before making a critical decision.

Mix of ML algorithms for spatio-temporal emission prediction: Emissions prediction by using spatial and temporal features is required to improve accuracy in DT-based transport management. As traditional ML methods are incapable of maintaining this [15], effective hybrid methods are required to achieve this and increase the accuracy of predicted values.

Design of an autonomous engine for sustainable and location-aware action recommendation: The next-generation transport networks require location-aware autonomous action recommendations for smooth and sustainable traffic flow. Such actions could be limiting the number and type of vehicles at the entrance of a junction during rush hours or putting speed limits at some proportion of the motorways.

Based on these motivations, our study moves along the research question, *"How to design a sustainable mobility framework in next-generation transportation networks (i) by precisely capturing the right-time data while preserving resources in querying, (ii) by concisely gaining insights on $CO2$ levels for forthcoming situations, (iii) by autonomously serving recommendations for the high-risk zones regarding the total number of vehicles and emission levels?"*. To address this, we propose the DT-empowered Eco-Regulation Framework with a novel twin sampling rate. In addition, we perform $CO2$ prediction within the service layer of DT to gain insight into future emission levels. Also, we perform risk zone clustering via Laplacian analysis to decide which locations need action recommendations to make them safe zones. After that, we perform Q-learning-based action recommendations to decrease emissions. As traffic dynamics can be unpredictable depending on the environment where multiple factors, like weather, accidents, or construction, affect the flow, we chose the Q-learning method to produce autonomous actions robustly.

TABLE 1. Proposed DTER Framework and Current State of the Art Studies

Literature	Data capturing	Emission prediction	Autonomy	Recomm. alg.	Context
[16], [17], [18], [19], [20], [21]	real-time	-	-	-	-
[22], [23], [24], [15]	historical	LSTM, CNN, ANN	-	-	-
[25], [26], [27], [28], [29]	real-time, historical	-	✓	RL	route planning
[30], [31] [32]	historical	-	✓	DRL, Deep-GAN	task scheduling, bus boarding
Our work	right-time	MLP, LSTM, Embedding	✓	Q-Learning	emission reduction

C. CONTRIBUTIONS

The contributions of this study are summarized below:

- We introduce a novel twin sampling rate formula to capture the right-time data according to the dynamics of transport networks. In the formulation, we consider incoming traffic flow, outgoing traffic flow, current traffic flow, and maximum flow metrics.
- We jointly use MLP, BANE embeddings, and Bidirectional LSTM to predict the future emission levels on a transportation network. With this, we serve the low-dimensional representation of spatial information of a transportation network within the twin service layer while increasing the prediction accuracy.
- We design an autonomous recommendation engine within the twin service layer to manage traffic flows by considering emissions. The engine is capable of Laplacian matrix-based risk zone clustering and Q-learning-based location-aware action recommendations.

II. RELATED WORKS

A. DATA CAPTURING EFFORTS IN DIGITAL TWINS

Despite the increasing studies in the DT research area, there is limited research on the main building block of DT, proper data capturing and modelling. In this regard, [16] studies the DT modelling and points out a data representation technique capable of holding all related information. The study also highlights that the flexibility of graphs and better querying performance may assist this tendency. Moreover, [17] presents a case study for the real-time situational analysis of the transportation system and maintaining informed traffic simulation models. The results guarantee the real-time delivery of traffic flow statistics and enhanced analysis. Likewise, [18] deals with the real-time data serving in the DT simulation of connected vehicles and pedestrians by proposing a closed-loop data transmission scheme. Furthermore, the use-case-oriented DT studies, [19], [20], mainly utilize Microsoft Digital Twin environment and utilized graph-based modelling. Furthermore, [21] aims to solve the timing problem in synchronising physical and virtual parts and computation tasks by minimizing the delay. For this, it proposes a mobile edge computing (MEC)-based load balance model to present the timely delivery of data within the DT. Even though these studies perform efficient data modelling schemes, none of them proposes specific data-capturing approaches to maintain right-time data within the DT system.

B. EMISSION PREDICTION MODELS IN TRANSPORTATION NETWORKS

In the current literature, multiple approaches exist for the management of emissions. In this regard, [22] proposes an eco-routing strategy by calculating the fuel consumption and, thus, resulting emission values with user equilibrium formulas. Even though the results contribute to decreasing the emissions significantly, the statistical method presented in this study cannot be applied directly to calculate future emission levels due to the dynamic nature of transportation networks. Moreover, [23] designs $CO2$ emission prediction methods for vehicles using data generated by in-vehicle sensors. They utilize LSTM models to capture the temporal dependencies in the time domain in a forward manner and predict future emissions. Likewise, [24] presents a $CO2$ modelling scheme by jointly using the physical and data-driven models. In this study, the physical model utilizes a cascaded architecture with the vehicles' features, while the data-driven models rely on a modified LSTM neural network. Similarly, [15] proposes a prediction system based on multiple 1D-Convolutional Neural Networks (CNN) and Artificial Neural Networks (ANNs), which predicts exhaust emissions based on speed, acceleration, and environmental parameters. As summarized above, there are several examples of emission prediction for transportation networks. Nevertheless, all of these studies consider either temporal features or spatial features, but they do not utilize them together within a DT system.

C. RECOMMENDATION FRAMEWORKS IN TRANSPORTATION DIGITAL TWINS

The primary focus in this area is vehicle routing strategies. For example, [25], [26] propose a route recommendation scheme that can timely produce output. However, in this proposed scheme, the emissions are not considered directly in the decision-making process, making them prone to favour the traffic flow. Moreover, [27] proposes a control strategy to mitigate congestion and lower emissions by actively controlling the lanes and speed limits. Nonetheless, this study utilizes a rule-based management system rather than an autonomous-driven one. In the study [28], an enhanced analysis of transportation systems is performed while proposing effective regulations for traffic flow management.

However, the decision-making process is not elaborated in detail, and the sole use of a deterministic simulation to generate actions may decrease the performance. Furthermore, [29] introduces an automatic transportation solution by proposing a Transportation Internet (TI) concept. In the TI, software-defined transportation (SDT) separates the control and transport planes. Even though this study presents a centralized intelligent control for transportation networks, emission-oriented recommendations are not addressed. Furthermore, the recommendation mechanism for task scheduling problems in transportation has been addressed with different methodologies, such as Deep Reinforcement Learning (DRL) [30], [31]. Recently, a generative adversarial network (Deep-GAN) is proposed for the bus boarding prediction and recommendation task [32]. Even though the results highlight improved performance and promising alignment with existing ridership information, this study does not target managing the emission levels. As a result, these studies primarily aim to improve traffic flow, with little consideration given to emissions in decision-making processes, possibly due to the limited resources and capabilities of traditional transport management systems.

We give the summary of the current state of the art in Table 1. According to this table, we note that the existing data capturing methods focus on real-time data and historical data. Our study is the first to propose right-time data capturing within a DT system for transportation networks. For emission prediction, the current studies perform traditional AI/ML algorithms with temporal features; they do not consider hybrid methods. To surpass this, we utilize hybrid methods with spatial and temporal features. Furthermore, we can also highlight that our study is the first to propose a traffic regulation scheme to decrease emissions. The remainder of the article is arranged as follows: Section III details the proposed Digital Twin-empowered regulation architecture. Section VI gives details about the experimental analysis. Lastly, the conclusion is given in Section V.

III. DIGITAL TWIN-EMPOWERED ECO-REGULATION FRAMEWORK

The proposed DT-empowered Eco-Regulation Framework is shown in Fig. 2. It comprises a Physical Twin Layer, a Digital Twin Layer, and a Twin Service Layer. In addition, the key notations are given in Table 2. Through our theoretical modelling and derivations, we use a bold font to show matrices.

A. PHYSICAL TWIN LAYER

This layer consists of the road infrastructure and vehicles. In this study, we discuss data-driven modelling, emission predictions and eco-regulations; therefore, we assume that any data regarding the traffic and infrastructure can be acquired. In this scope, the measured data on the junctions regarding the traffic flow, such as incoming flow value to the junction and outgoing flow value from the junction, is utilized. More specifically, for each vehicle travelling on the road, we take the source junction road name, destination junction road name, record timestamp, and direction of travel information. In addition, we take latitude and longitude records to use in the spatial analysis.

TABLE 2. Nomenclature

Notation	*Explanation*
G	Knowledge graph of transport topology
G'	Predicted graph of transport topology
$\mathcal{G}$	Base knowledge graph
M_S	Spatial attributes
M_T	Temporal attributes
X'	Predicted $CO2$ signals
B	Embedding Matrix
Z_i	i^{th} risk zone
e_i	Average $CO2$ value in the i^{th} risk zone
n	Number of nodes in graph G'
r	Total number of risk zones in G'
k	Total number of nodes in the i^{th} zone
$\hat{\alpha}$	Twin sampling rate
ρ_i	Utilization of the node i
ζ_i	Dynamicity factor for node i
C_i	Maximum current flow for node i

B. DIGITAL TWIN LAYER

1) Right-time Data Capturing

We introduce a twin sampling rate specific to the nature of a transport topology to hit the dynamical changes and ensure physical-to-virtual convergence. We explain the details of this metric below.

Twin Sampling Rate, $\hat{\alpha}$: It refers to the synchronisation frequency between the physical object and its digital counterpart. Therefore, we define the twin sampling rate for a transport network to ensure physical-to-virtual convergence while catching the dynamic flow variations and avoiding loss of information. We first define the utilization and dynamicity factors for a junction in a transport network. We calculate the utilization (ρ) of a junction i in a transportation network as:

$$\rho_i = \frac{flowCur_i + flowIn_i - flowOut_i}{C_i} \ , \ \forall i \qquad (1)$$

where, $\rho_i \in [0, 1]$. In (1), we consider the current flow value within the junction $flowCur$, the incoming flow value to the junction $flowIn$, and the outgoing flow value from the junction $flowOut$. All these flow values are in $vehicles/min$. Also, the term C_i stands for the maximum current flow of the junction i for vehicles. Then, we define the dynamicity factor (ζ) to measure the traffic flow changes. This is because we need a trigger from the physical environment to decide the synchronization time, which depends on the behavioural pattern of the transport topology. As the

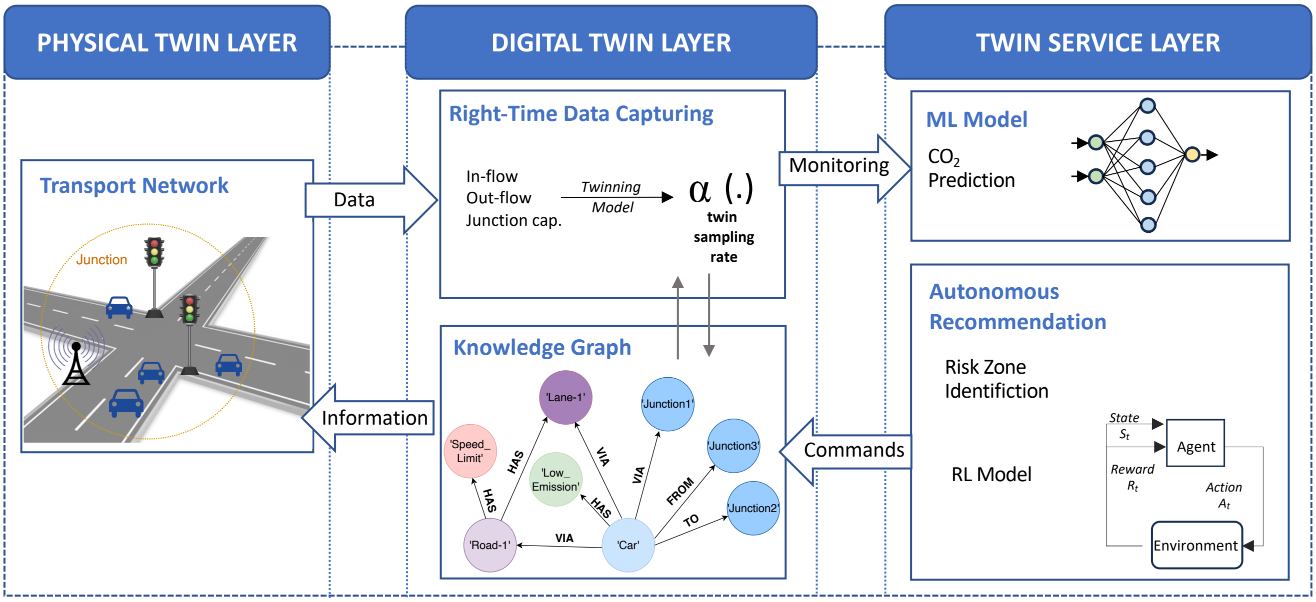


FIGURE 2. **DT-empowered Eco-Regulation (DTER) framework.**

main dynamicity stems from the traffic flow, we form the dynamicity factor of a junction i as given below:

$$\zeta_i = \frac{flowIn_i + flowOut_i + |\Delta flowCur_i|}{2(C_i + flowIn_i)}\ , \ \forall i \qquad (2)$$

where, $\zeta_i \in [0, 1]$. In (2), we consider the metrics appertaining to the traffic change. Regarding this, we use incoming traffic flow, outgoing traffic flow, and maximum current flow as we did in the utilization. In addition, we use the current traffic flow change that is $\Delta flowCur_i$, implying the total change in the traffic flow when compared to the previous time step. When forming (1), and (2) we use two upper bounds:

- *Upper bound-1:* This is to ensure both the incoming flow value to a junction and the current flow value within the junction should be less than the maximum current flow of the junction. We denote this bound as $flowIn, flowCur \leq C$. The main reason the incoming flow value to a junction should be less than the maximum current flow is to avoid additional congestion occurrences.
- *Upper bound-2:* This bound ensures that the outgoing flow value cannot exceed maximum current flow and the incoming flow to a junction. We denote this bound as $flowOut \leq C + flowIn$. This is a kind of application of the flow conservation rule; if there is no incoming flow and current flow in a junction, then there will be no outgoing flow.

In the calculation of (1), if we take the maximum values according to these bounds, we see that the numerator will be equal to $C - flowIn$. And if this value is divided by C, it will result in a value lying in $[0, 1]$. By applying these bounds, we ensure the modelled transport system's stability by setting limits according to the maximum current flow values of the different junctions. After that, we use the junction-specific utilization and dynamicity factors in order to decide the network's overall twin sampling rate. We calculate the twin sampling rate as given below:

$$\hat{\alpha} \approx \max\{\zeta_i \rho_i | i = 1, 2, ...n\} \qquad (3)$$

In this formula, $\zeta_i \rho_i$ values stand for the individual twin sampling rates of the junctions. Here, the utilization factor implies the importance of a junction in a transportation network, while the dynamicity factor highlights the flow changes for a junction. As we want to catch the dynamic changes for the important junctions of a transportation network, we multiply these two metrics to find the individual twin sampling rates of the junctions. Similarly, to catch the dynamical changes in the transport topology and update our records accordingly, we take the maximum twinning rate among the set of junctions to apply for the whole transportation topology. In this way, we can hit the required twinning rate for the most dynamic scenario by involving the less dynamic ones as well. With the result of (3), we fed the digital twin layer with the decided $\hat{\alpha}$ and updated the knowledge graph, G, accordingly. According to $\hat{\alpha}$, if a node's traffic change is high, the twin sampling rate value will have a high value and vice versa. The practical meaning of this rate is that it decides when to update the data in the knowledge graph by perceiving the traffic flow information from junctions.

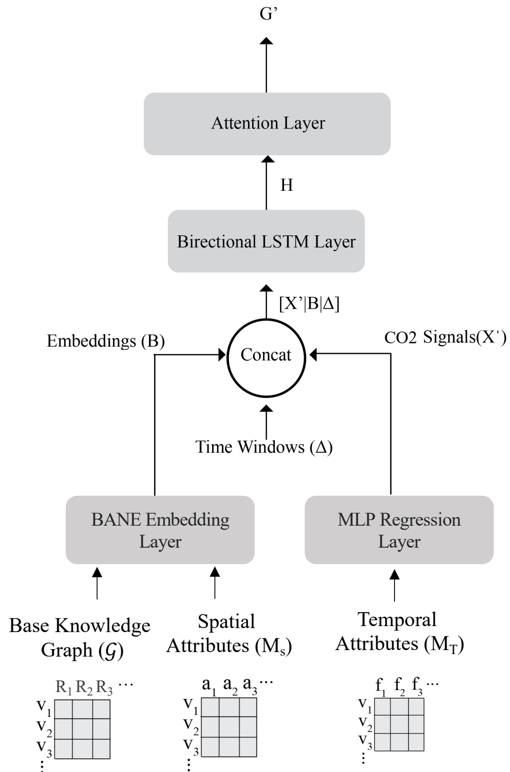


FIGURE 3. $CO2$ **Prediction in Twin Service Layer.**

2) Knowledge Graph Construction

The modelling methodology for transportation networks is essential to create a 360-degree view. To do this, the real-time acquired data should be represented with additional information, such as road infrastructure, relations between transport entities, and regulation rules. In addition, these information sets need to be available within the transportation network model with spatial and temporal indexes. To enable such an environment, we first start with constructing the knowledge graph of the transport network. We represent this knowledge graph as $G = (V, E, R, M_s, M_T)$, where V, E, and R represent the node set, edge set, and regulation set of the transport topology, respectively. The nodes represent the intersections of nodes, while the edges correspond to their contextual and spatial relationships. The regulations imply the set of rules to be applied for corresponding junctions. In addition, M_S and M_T correspond to spatial and temporal attributes of the knowledge graph that reflect the specific information of the given point, such as GPS location, speed limit, road direction, and $CO2$ level. At this point, we call a subgraph of G, as base knowledge graph to utilize in $CO2$ prediction in the next section and we denote it as $\mathcal{G} = (V, E, R)$, where $\mathcal{G} \subset G$ is hold.

C. TWIN SERVICE LAYER

1) $CO2$ Prediction

DTER utilizes a hybrid method for emission prediction. More specifically, it comprises four modules: BANE Embedding Layer, Multi-Layer Perceptron (MLP) Regression Layer, Bidirectional LSTM Layer, and Attention Layer. The proposed $CO2$ prediction scheme is illustrated in Fig. 3. Here, the machine receives the base knowledge graph ($\mathcal{G}$), spatial attributes (M_S), temporal attributes (M_T), and time window (Δ). The Δ vector is the corresponding time difference between each timestep.

BANE Embedding Layer: Node embeddings have been used to represent nodes in a graph as low-dimensional vectors while preserving the structural properties and relationships. These are generated by algorithms such as random walks, singular value decomposition, and graph neural networks. These plain network embeddings can fall short of representing real-world conditions because of a lack of considering attributes such as occupancy rate and travel time variability. Thus, we utilize the attributed graph embedding algorithm, BANE model [33], to preserve structural and contextual information. This BANE embedding layer takes in $\mathcal{G}$ and M_S and creates binary codes for vertices, resulting in the embedding matrix $\mathbf{B}$.

MLP Regression Layer: The junctions' incoming and outgoing flow data are contained within the temporal attributes $M_T \in R^{N \times F \times T}$. Using this, the MLP regression module calculates the produced $CO2$ emissions. This module receives $N \times F$ inputs and generates N outputs, creating $X' \in R^{N \times T}$.

Bidirectional LSTM Layer: The emission levels throughout the time domain might follow certain patterns resulting from the traffic patterns. The Bidirectional LSTM considers long- and short-term patterns in both forward and backward directions. Here, this layer receives the concatenation of node embeddings (B), $CO2$ signals(X'), and time window (Δ). Here the X' is the matrix containing the predicted emissions per node. The hidden state matrix (H) of the Bidirectional LSTM is forwarded to the attention layer.

Attention Layer: The attention mechanism computes a weighted sum in which the weights are learned. This layer calculates the resulting degree of emission considering the respective timesteps. Using this layer, the model can consider both $CO2$ production and its accumulation. Then, this model outputs the emission predictions, denoted as G'.

2) Autonomous Action Recommendation

The proposed autonomous recommendation engine comprises two significant steps. First, it performs risk zone clustering for detecting the high-risky zones in terms of $CO2$ levels. Here, the risk zones represent particular areas within the transport topology having differentiated $CO2$ values. Then, a Reinforcement Learning (RL)–based action set is explored. The details of these steps are explained below.

Risk Zone Clustering: The $CO2$ values in a transport network depend on the dynamics of the flowing traffic, and the level of emissions differ spatially depending on these dynamics. Therefore, performing clustering for risk zones

This article has been accepted for publication in IEEE Open Journal of Vehicular Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/OJVT.2024.3484956

enables us to understand emission behaviour by catching the zone-specific emission values and producing accurate recommendations accordingly. At this point, we formulate clustering as a community detection problem in graph theory. To solve this emissions' community detection, we work on the predicted graph, $G^{'}$, and refer to each community as a separate risk zone, Z_i. Therefore, our approach holds the statement $G' = \bigcup_{i=1}^{r} Z_i,\ G' = Z_1 \cup\ Z_2 \ldots Z_r$ where r is the total number of risk zones within the transport topology.

$CO2$ **Matrix, O**: It is an $n \times n$ matrix comprising of the $CO2$ values for individual nodes in graph G'. We represent it as **O** = $diag(o_1, ..., o_n) \in R^{nxn}$. The diagonal elements of this matrix stand for the node-specific predicted $CO2$ values. Therefore, the non-diagonal elements are equal to zero.

$$\mathbf{O} = \begin{bmatrix} o_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & o_n \end{bmatrix}$$

Strength Matrix, S: It is an $n \times n$ matrix stating the strengths between the pair of nodes in a neighbourhood area within the transport topology. The strength of a node pair is determined by the total edge weights connecting them. As in our implementation, we denote the junctions as nodes and the roads as edges; the edge weights are proportional to the physical distance between the two junctions. We represent it as **S**= $\{s_{ij}|i,j = 1,2,\ldots n\}$, where s_{ij} indicates the strength value of node pairs (i, j). The diagonal elements of this matrix are equal to zero implying that self-loops will contribute to zero distance.

$$\mathbf{S} = \begin{bmatrix} 0 & s_{12} & \ldots & s_{1n} \\ s_{21} & 0 & \ldots & s_{2n} \\ \vdots & s_{(n-1)2} & 0 & s_{(n-1)n} \\ s_{n1} & s_{n2} & \ldots & 0 \end{bmatrix}$$

Laplacian matrix analysis serves to preserve the local geometry, which is the case we desire for our transport network implementation. Therefore, we construct the Laplacian matrix of our transport network for the zone identification task. With this, the nearby points in the physical space will remain nearby in the reduced virtual space. Regarding this, there are different ways of defining the Laplacian matrix of a graph depending on the application. As our main target is to create clusters on our knowledge graph, we use the Normalized Laplacian matrix and one of its properties. We use (4) to generate the Normalized Laplacian matrix, which partitions the network into clusters based on the smallest eigenvalues and their corresponding eigenvectors.

$$\mathbf{L}^N = 1 - \mathbf{SO}^{-1} = \begin{cases} 1, & i = j \\ -s_{ij}/o_j, & (i,j) \in G \\ 0, & o.w. \end{cases} \tag{4}$$

Moreover, we partition the graph into clusters by using the smallest eigenvalues and corresponding eigenvectors by using Spectral Clustering with the k-means algorithm. We also utilize a stopping criteria for the clustering, as explained below.

Community Matrix, Z: It states the node-specific $CO2$ values and highlights the clusters in the graph as a result of Laplacian matrix analysis. For instance, the community matrix of a graph network with two extracted clusters will look like:

$$Z = \begin{bmatrix} z_{11} & \cdots & z_{1m} & \cdots & z_{1n} \\ z_{21} & \cdots & z_{2m} & \cdots & z_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ z_{m1} & \cdots & z_{mm} & \cdots & \cdots \\ z_{(m+1)1} & \cdots & z_{(m+1)m} & \cdots & z_{(m+1)n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ z_{n1} & \cdots & z_{nm} & \cdots & z_{nn} \end{bmatrix}$$

As we desire to form clusters regarding the $CO2$ values, we calculate the average emission value in percentage (e_i) and take it as a stopping criterion for iterative clustering. We calculate e_i value for the cluster i as given in (5):

$$e_i(\%) = \frac{1}{k} \sum_{j=1}^{k} z_{jj} \times 100 \tag{5}$$

where the value of k corresponds to the total number of nodes present in cluster i. Therefore, $k \leq n$ should be held.

Q-learning based Recommendation Engine: Since Q-learning works on discrete states and actions, it fits well for our target, which is making location-aware recommendations to reduce emission values. That's why we adapt Q-Learning for the autonomous creation of actions by deciding the number of vehicles to be accepted for a particular junction. Therefore, we map the Q-learning components into our transport topology by considering each node, namely junctions, as an agent. Also, we consider each cluster calculated with the Laplacian as a state. The agent performs action by exploring the number of vehicles to be served within a junction and obtains a reward for this. Therefore, our recommendation system is represented as six tuples; $\{E, Ag, S, A, p, R\}$. We explain the major elements of the recommendation system below.

- *Environment, E:* The transport network consisting of n number of junctions forms an environment for Q-learning. The transport topology is constructed in the form of $CO2$ zones, as explained in the above section.
- *Agent, Ag:* Each junction in our transport network that can calculate the in-flow and outflow traffic is defined as an agent.
- *State, S*: The state space represents the total number of $CO2$ classes consisting of three types as quantized in Table 3. We define these classes' boundaries in the **Z** by referencing the e_i values. This is because average

emission values differ depending on the dynamics of transport topology.

- *Action, A:* Action set states the total number of vehicles to be served at a time while keeping the $CO2$ values below the risky level. These risky situations and action space values are stored within the knowledge graph, G. Here, the $action_i$ is formed for Z_i where $i = argmax(e_i)$. Then, the action set is formed as $A = \{action_i | i = 1, 2, ...r\}$ for each of the risk zones.
- *Probability, p:* Represents the probability of transition to the new state. In our approach, it stands for the probability of transitioning from one emission class to another. Therefore, this value is important while defining the boundaries of threshold values.
- *Reward, R:* The agents accept a reward for each emission class change in which the emission value is decreased, considering the transport topology. In this circumstance, we form an emission-oriented reward function for this study. To do this, we define an emission function to be minimized and refer to it while creating the reward function. Therefore, we first calculate the geodesic distance, which is the shortest path between the two nodes, by using the betweenness centrality measure. We choose this centrality as it considers the location information of a node with the global view of the entire transport network. In addition, we take advantage of its flow consideration perspective, which is impossible with other centrality measures. We calculate the dynamical betweenness centrality value of a node as follows: $b_i = \frac{1}{(n)(n-1)} \sum_{l \neq i, m \neq i, l \neq m} \frac{\rho_{lm}(i)}{\rho_{lm}}$ [34]. As the betweenness centrality value increases with the number of nodes in a graph, we use the relative betweenness centrality value that we calculate by using normalization: $b'_i = \frac{b_i}{MAX(b_i)}$ to hold $0 \leq b'_i \leq 1$ condition. We use the b'_i values in our objective function to decrease the $CO2$ values considering the graph G'. After that, we form the *Emission Function* denoted as E, as given in (6). This function sums the total cluster-specific calculated emission values by considering the centrality measure and the average emission value within the cluster. The constraint (7) ensures the total number of clusters should not exceed the total number of nodes in transport topology. Moreover, the centrality value ($b_i^{'}$) should be in the interval $[0, 1]$ (8). Lastly, the average emission values (e_i) within a cluster should not be non-negative for all updated adaptive twinning rates (9).

$$\text{Minimize} \quad E = \sum_{Z_i} b_i^{'} e_i \qquad \forall Z_i \in G' \tag{6}$$

$$\text{subject to} \quad \sum_{i \in V} Z_i \leq n \quad \forall i = 1, m \tag{7}$$

$$0 \leq b_i^{'} \leq 1 \quad \forall n \in V \tag{8}$$

$$e_i \geq 0 \quad \forall \hat{\alpha} \tag{9}$$

Then, we form an emission-oriented reward function:

$$R_{t+1} = \frac{(e_t - e_{t+1})^+}{E} \tag{10}$$

where $(y)^+$ operator takes the value of y if it is positive, and takes zero otherwise. In (10), the numerator checks for if the action leads to a decrease in the $CO2$ level. Here, dividing it by the total emission of the whole topology, E, gives the normalized reward value. At this point, we note that DTER focuses on reducing road transportation emissions. More specifically, if a congestion scenario is to be avoided, additional route recommendation algorithms should be integrated into the system. Thus, in this study, we assume that DTER does not result in any congestion or accident scenario.

- *Q Function:* The algorithm updates the Q values by using the equation which we adapt from Bellman Equation: $Q_{t+1}(S, A) = Q(.) + \alpha(R + \gamma \max_{A'} Q(S', A') - Q(.))$, where $Q(.)$ function stands for $Q_t(S, A)$. In this formula, the agent, a junction, considers all the possible actions, that is, the number of vehicles and states of transport regions, to choose one, maximizing the reward function. In this formula, the A' is the action that could be taken at state S'. Also, γ is a discount factor showing the significance of the next states with learning rate, α.
- *Policy, π:* We utilize the ϵ-greedy action selection mechanism by randomly choosing the exploration and exploitation states. Therefore, the junction agent takes a random action at a given time with the probability of ϵ or $(1 - \epsilon)$.

TABLE 3. Emission Classes and Corresponding Boundaries

Emission Class	Threshold
1	$e_i \leq 30\%$
2	$30\% < e_i \leq 70\%$
3	$e_i > 70\%$

As seen in Alg.1, DTER takes the spatio-temporal features and Q-learning parameters information as the inputs and produces the base knowledge graph and whole knowledge of the transport topology (line-2 in Alg.1). After that, it performs $CO2$ prediction steps and constructs the predicted knowledge graph by including the emission values for each junction (line-7 in Alg.1). Then, transport specific twin sampling rate is calculated by considering the utilization and the dynamicity factors of the related junction (line-9 in Alg.1). The clustering stopping criteria is checked in line-10. If the stopping condition is not met, Laplacian matrix analysis is performed in line-12 and line-13. If the condition is met, the clustering is stopped, and the cluster information is used as in the last updated community matrix. The emission function value is calculated in line-19 with the given constraints. After all these steps, Q-learning is performed by taking an action with the ϵ-greedy approach and the Q-table is updated according to the reward function

This article has been accepted for publication in IEEE Open Journal of Vehicular Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/OJVT.2024.3484956

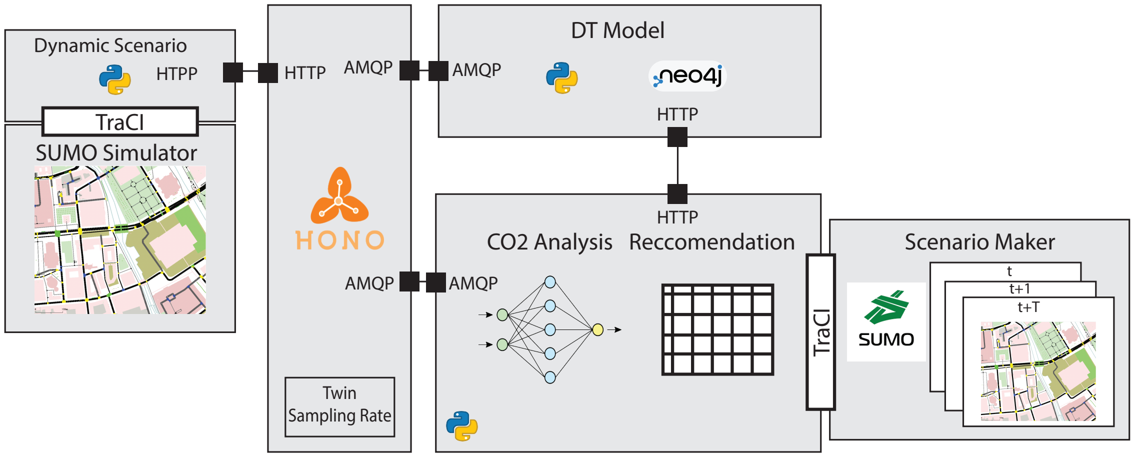


FIGURE 4. Simulation environment.

Algorithm 1 DTER: DT-empowered Eco-Regulation Algorithm

Require: spatio-temporal feature set, α, γ
Ensure: G, G', $\mathcal{G}$, Z , E, R, A_{t+1}

1: Initialize $\hat{\alpha}$, M_S, M_T, e_i, Q, A;
2: Construct G, and $\mathcal{G}$;
3: **foreach** change in $\hat{\alpha}$
4: Update M_S, and M_T
5: Perform MLP to create X'
6: Retrieve **B**
7: Perform LSTM to create G'
8: Calculate ρ and dynamicity factor ζ ▷ (1),(2)
9: Calculate $\hat{\alpha} \approx \max \{\zeta_i \rho_i | i = 1, 2, ...n\}$ ▷ (3)
10: **while** $|c_j - c_k| > e_i$
11: Construct **O, S**
12: Generate L^N, Compute eigenvectors ▷ (4)
13: Apply k-means, Generate **Z**
14: Calculate $e_i = \frac{1}{k} \sum_{j=1}^{k} z_{jj} \times 100$ ▷ (5)
15: **end**
16: **end**
17: Calculate centrality, $b_i = \frac{1}{(n)(n-1)} \sum_{l \neq i, m \neq i, l \neq m} \frac{\rho_{lm}(i)}{\rho_{lm}}$
18: Calculate $b_i' = \frac{b_i}{MAX(b_i)}$
19: Calculate emission function, E ▷ (6)
20: **foreach** episode
21: Take action, a
22: Observe reward R, and state, S' ▷ (10)
23: Update Q-table
24: $Q_{t+1} \leftarrow Q(.) + \alpha(R + \gamma \max_{A'} Q' - Q(.))$
25: Decide next action, a' with policy, π
26: Update $a \leftarrow a'$, $S \leftarrow S'$
27: Update action set, $A_{t+1} \leftarrow A_t$
28: **end**

results (lines 20-26). In the last step, the optimum action set is applied as the output of the autonomous recommendation engine (line-27).

IV. EXPERIMENTAL STUDIES

A. DATASET

Due to the scarcity of reliable on-road carbon emission datasets, we conduct experiments by using a real-world dataset from the open-source project[1]. We first create our physical twin layer in SUMO by extracting the traffic volumes in the UK from four major roads: A330, A3095, A329, and A321. For this, we used the data from the roads dataset and merged them with the count points and count entries dataset. The raw data we use in our simulations includes 1000 transactions and 280 unique trips, each with different source and destination junctions. Afterwards, we create two scenarios for emission values:

- *Scenario-1:* In this, we create emission values of the vehicle records by utilizing the predefined emission models in SUMO. We set the simulation area as 1 km^2. We assume that all the vehicles are identical. Therefore, we present the HBEFA3/PC G EU4 model for Euro norm 4 gasoline cars to record emission values.
- *Scenario-2:* This is to test our proposed scheme on a more realistic scenario with real emission values. For this, we extract the emission values from the American on-road carbon emission database, DARTE [35], especially 2017 records with the corresponding latitude and longitude information.

B. SIMULATION ENVIRONMENT

The simulation environment is shown in Fig. 4. Here, the Physical Twin Layer is constructed using the microscopic and continuous traffic simulator Simulation of Urban Mobility (SUMO) [36] where we mimic the real-world transportation network. Moreover, a dynamic scenario module is developed using Python and integrated with the SUMO simulator using the TraCI interface. Moreover, the connection between the simulation environment is established using the Eclipse

[1] https://github.com/Software-Dev-Group-Project/traffic-analyzer-sdgp

Hono, a cloud-based IoT device management platform[2]. To communicate with Hono, we used the HTTP protocol. Here, the data measurements are sent to, and control commands are received from the Hono. The integration of Hono to the DT model and Service Layer applications is established using the AMQP messaging protocol. Also, the updates to the DT model hosted in Neo4j[3] and the data handling and service applications of the proposed architecture are realized by Python scripts. We used the HTTP protocol between the twin service layer and the DT model. Additionally, the Scenario Maker module is implemented with SUMO. The traffic demand, topology, and action set are received to set the environment. The evaluation, monitoring and control of this environment are established via the TraCI interface. All simulation parameters are given in Table 4.

C. PERFORMANCE RESULTS

In this section, we aim to investigate the performance of DTER considering (i) the right-time data capturing with the proposed twin sampling rate and also mean query duration to access an increasing number of objects in the knowledge graph, (ii) the accuracy of the prediction algorithms implemented within the Twin Service Layer, and (iii) total $CO2$ emission reduction within the zones via Q-learning based autonomous actions and update mechanism (in g/km^2) *for Scenario-1 and Scenario-2.*

TABLE 4. DTER Network Simulation Parameters

Parameters	Values
Topology Parameters	**Total Number of Nodes:** {25, 50, 250} **Incoming traffic flow:** [0, 100] veh/min (identical vehicles) **Max. current flow:** 100 veh/min
ML Model Parameters	**LSTM:** **Optimizer:** adam **Merge mode:** concat (for BiLSTM) **Recurrent initializer:** orthogonal **Max epochs:** 40 **Batch size:** 256 **MLP 2 (5,2)** **Optimizer:** sgdm **Cross-validation:** 4-fold **Max epochs:** 40 **Batch size:** 256
RL Model Parameters	**Q-Learning:** **Num. of episodes:** 10000 **Discount factor** 0.99 **Learning rate:** 0.1 **Update policy:** Epsilon-greedy

We firstly search for the effect of twin sampling rate, $\hat{\alpha}$, on the right-time data capturing for the simulated transport

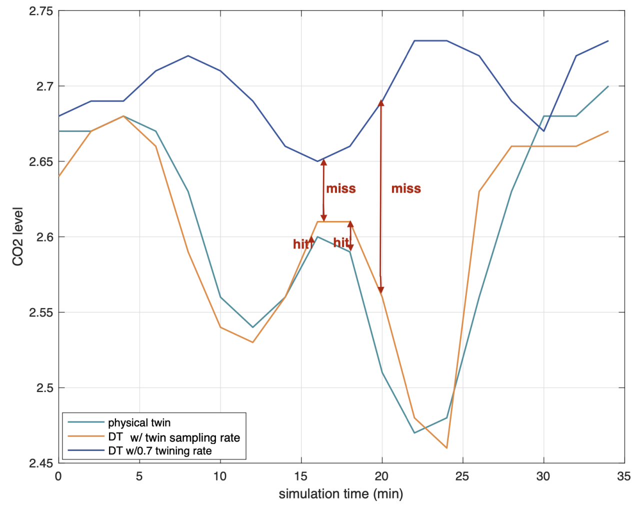


FIGURE 5. Digital Twin Layer performance analysis in terms of right-time data capturing for proposed twin sampling rate and constant twinning rate.

network within the Digital Twin Layer. For this, we create a time series-based dynamic road network scenario in a Python script with different incoming flow, outgoing flow, and current flow for each junction. We also set maximum current flow value as given in Table 4 for each junction. In addition, we graduate $CO2$ values into levels in the dynamic scenario and work on the interval of [2,3]. These $CO2$ levels correspond to the physical twin emission values and are represented in Fig. 5 with the light blue line. We use this as a baseline to compare the results of the proposed twin sampling rate and a constant twinning rate of 0.7. We simulate for thirty-five minutes and observe the resulting $CO2$ levels for these twinning rates. We labelled the cases as $miss$ or hit. A $miss$ case occurs when the resulting $CO2$ level differs at a lower bound of 15% from the physical twin emission level. On the contrary, a hit case implies an upper bound of 5% to be accepted as successive modelling for the dynamic scenario. Regarding these limits, we have observed the $miss$-hit cases during the simulation time. The results show that the proposed twin sampling rate gives 19% more successful right-time data capturing than the constant 0.7 twinning rates. More specifically, we see that the proposed twin sampling rate surpasses the constant rate method when catching the traffic flow behaviour of the dynamic scenario due to its continuous trace on traffic flow metrics. The results for the entire simulation time are given below.

Moreover, we observe the required duration for a query to test the performance of the implemented knowledge graph hosted in Neo4j within DTER. For this, we have increased the number of requested objects to one thousand in a single query and noted the required duration in ms. We individually compare the performance of the knowledge graph with traditional data files and relational databases. As the number of requested objects increases, traditional and relational databases take more time to retrieve all the

[2]https://www.eclipse.org/hono/
[3]https://neo4j.com/

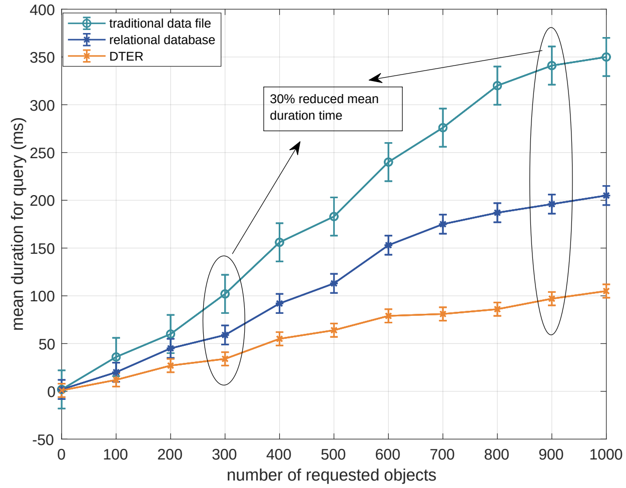


FIGURE 6. **Digital Twin Layer performance analysis in terms of mean duration performance to reach the requested objects for traditional data file, relational database, and knowledge graph.**

requested objects due to intensive join operations performed on the data files. This can be seen in Fig. 6. On the other hand, as our proposed DT modelling takes advantage of preserving relationships while separating the real-time data into spatio-temporal graphs, we see a 30% reduced mean duration time to reach all requested objects compared to the two circles in the figure.

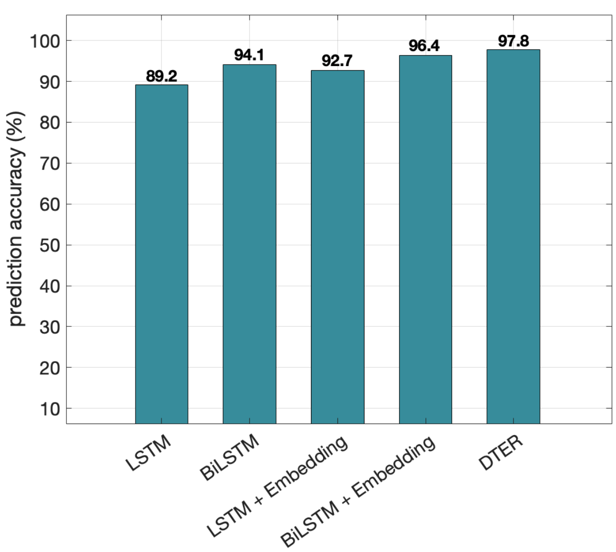


FIGURE 7. **Twin Service Layer performance analysis for $CO2$ prediction for LSTM, BiLSTM, embeddings integrated, and DTER.**

In addition, we test the DTER performance regarding the $CO2$ prediction within the Twin Service Layer. For this, we decided on a 2-layered 5-2 architecture for MLP to estimate the current emission based on the graph signals of the knowledge graph. In the proposed method, we use $concat$ as the merge mode of the graph embeddings before giving it to BilSTM. We compare our prediction method with a single implementation of LSTM and BiLSTM and the joint usage of these with the graph embeddings. As seen in Fig. 7, the most accurate prediction arises with our proposed model with 97.8% accuracy, where we add MLP to the joint usage of BiLSTM and embeddings.

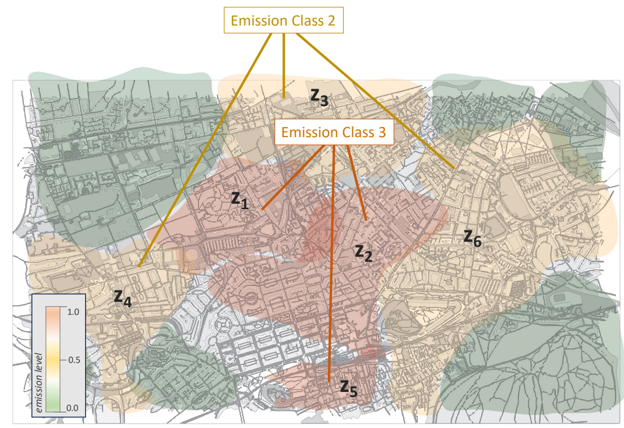


FIGURE 8. **Extracted emission classes.**

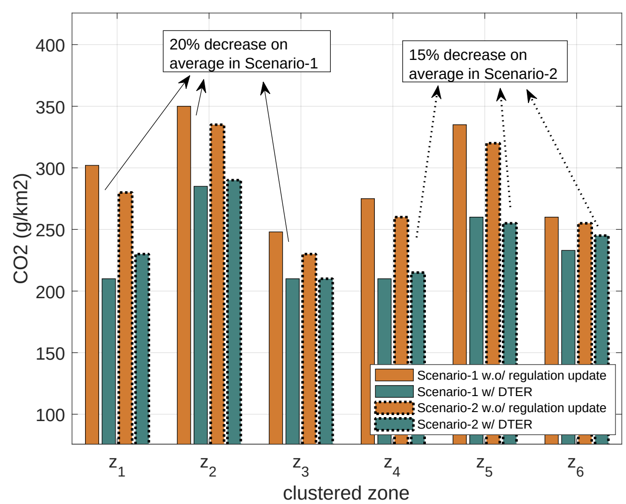


FIGURE 9. **$CO2$ reduction with DTER recommendation engine for Scenario-1 and Scenario-2.**

In the last section, we search for the effect of proposed regulation updates within DTER on the total $CO2$ emissions for the clustered zones individually. For this, we integrate our dynamic scenario into the SUMO environment. We run *Scenario-1* and *Scenario-2* separately. We set the same maximum current flow values for both of the scenarios. Similarly, we set the maximum allowable traffic flow density at 3 veh/m^2 to avoid additional congestion occurrences. In the What-If implementation, we define our objective function and initialize the main parameters (given in Alg.1) for the first run of the Q-learning. We compare the performance of DTER with the no-action applied case by considering the two scenarios. Fig. 8 shows the result of Laplacian matrix analysis for *Scenario-1* presenting six emission classes in total. Three of these classes belong to the Emission Class-3 due to their average emission value having the value of $e_i > 70\%$, while the remaining three of them lying within $30\% < e_i \leq 70\%$. We use these clustered zones in the x-axis of Fig. 9 while evaluating the average emission values with DTER and without applying the regulation updates. As seen in this figure, in *Scenario-1*, when no regulations are applied, the total $CO2$ value reaches the above three hundred levels. Conversely, with DTER, the $CO2$ values are controlled via the autonomous recommendation of the number of vehicles allowed for junctions. This control results in $\sim 20\%$ reduction in the total emission values. Similarly, in *Scenario-2*, we record a 15% reduction in emission levels with DTER. The main reason for this reduction stems from a higher miss rate when we embed real-world emission data into the SUMO environment. As a result, we can analyze the forthcoming emission values by changing the scenario parameters and acting accordingly before a condition occurs in the real environment.

V. CONCLUSION AND FUTURE WORK

This study presents a DT-empowered Eco-Regulation framework (DTER) to design next-generation transportation networks while addressing mobility effectively. For this, DTER addresses the challenges of implementing DT within transportation systems. Firstly, it fills the gap in the literature regarding right-time data capturing by presenting a twin sampling rate specific to transport network dynamics. With this, the proposed scheme successfully captures the right-time data by 19% more successively and decreases the query duration by 30% via the knowledge graph utilization. In addition, the accurate $CO2$ prediction mechanism with MLP, Bidirectional LSTM, and node embeddings makes DTER a strong method to infer future emission cases. In addition, DTER performs Laplacian-based emission zone clustering to highlight the risky locations. Lastly, the autonomous recommendations of DTER consider the predicted emission values and contribute to decreased emission levels by $\sim 20\%$ with SUMO emission models, and by 15% with real-world emission values.

For future work, we plan to improve our twin sampling rate formula by differentiating the several types of vehicles and their corresponding emission values. Moreover, we want to explore GAN methods for regulation formation. This will allow us to enhance the capabilities of our proposed model and test it from scalability and robustness perspectives.

ACKNOWLEDGMENT

This work was supported by The Scientific and Technological Research Council of Turkey (TUBITAK) 1515 Frontier R&D Laboratories Support Program for BTS Advanced AI Hub: BTS Autonomous Networks and Data Innovation Lab. Project 5239903. The work of T. Q. Duong was supported in part by the Canada Excellence Research Chair (CERC) Program CERC-2022-00109.

REFERENCES

[1] Z. Tu, L. Qiao, R. Nowak, H. Lv, and Z. Lv, "Digital twins-based automated pilot for energy-efficiency assessment of intelligent transportation infrastructure," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 11, pp. 22 320–22 330, 2022.

[2] N. S. Rajput, U. Singh, A. Dua, N. Kumar, J. J. P. C. Rodrigues, S. Sisodia, M. Elhoseny, and Y. Lakys, "Amalgamating vehicular networks with vehicular clouds, ai, and big data for next-generation its services," *IEEE Transactions on Intelligent Transportation Systems*, vol. 25, no. 1, pp. 869–883, 2024.

[3] Department for Business, Energy & Industrial Strategy, "2019 uk greenhouse gas emissions," 2019, [Accessed: 10-Mar-2023]. [Online]. Available: https://www.gov.uk/

[4] J. Gao and L. Zhen, "Intelligent distribution transportation route planning considering traffic congestion," *IEEE Intelligent Transportation Systems Magazine*, vol. 16, no. 3, pp. 96–110, 2024.

[5] M. Chae, T. Kim, and D. Won, "Optimal operation scheme of electric vehicle routing and charging considering power distribution and transportation integrated system," *IEEE Access*, vol. 12, pp. 83 427–83 438, 2024.

[6] F. Chen, Y. Lu, L. Liu, and Q. Zhu, "Route optimization of customized buses based on optimistic and pessimistic values," *IEEE Access*, vol. 11, pp. 11 016–11 023, 2023.

[7] X. Zhang, P. Wu, C. Chu, and M. Zhou, "A distributionally robust optimization for reliability-based lane reservation and route design under uncertainty," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 12, pp. 14 490–14 505, 2023.

[8] Y. Wu, K. Zhang, and Y. Zhang, "Digital twin networks: A survey," *IEEE Internet of Things Journal*, vol. 8, no. 18, pp. 13 789–13 804, 2021.

[9] M. S. Irfan, S. Dasgupta, and M. Rahman, "Toward transportation digital twin systems for traffic safety and mobility: A review," *IEEE Internet of Things Journal*, vol. 11, no. 14, pp. 24 581–24 603, 2024.

[10] M. Jafari, A. Kavousi-Fard, T. Chen, and M. Karimi, "A review on digital twin technology in smart grid, transportation system and smart city: Challenges and future," *IEEE Access*, vol. 11, pp. 17 471–17 484, 2023.

[11] H. Yan, X. Xu, M. Bilal, X. Xia, W. Dou, and H. Wang, "Customer centric service caching for intelligent cyber–physical transportation systems with cloud–edge computing leveraging digital twins," *IEEE Transactions on Consumer Electronics*, vol. 70, no. 1, pp. 1787–1797, 2024.

[12] L. Nie, X. Wang, Q. Zhao, Z. Shang, L. Feng, and G. Li, "Digital twin for transportation big data: A reinforcement learning-based network traffic prediction approach," *IEEE Transactions on Intelligent Transportation Systems*, vol. 25, no. 1, pp. 896–906, 2024.

[13] Y. Tian, A. Carballo, R. Li, and K. Takeda, "Rsg-gcn: Predicting semantic relationships in urban traffic scene with map geometric prior," *IEEE Open Journal of Intelligent Transportation Systems*, vol. 4, pp. 244–260, 2023.

[14] K. Duran, M. Özdem, T. Hoang, T. Q. Duong, and B. Canberk, "Age of twin (aot): A new digital twin qualifier for 6g ecosystem," *IEEE Internet of Things Magazine*, vol. 6, no. 4, pp. 138–143, 2023.

[15] X. Fei, F. Long, F. Li, and Q. Ling, "Multi-component fusion temporal networks to predict vehicle exhaust based on remote monitoring data," *IEEE Access*, vol. 9, pp. 42 358–42 369, 2021.

[16] L. V. Cakir, K. Duran, C. Thomson, M. Broadbent, and B. Canberk, "Ai in energy digital twining: A reinforcement learning-based adaptive digital twin model for green cities," in *ICC 2024 - IEEE International Conference on Communications*, 2024, pp. 4767–4772.

[17] H. Xu, A. Berres, S. B. Yoginath, H. Sorensen, P. J. Nugent, J. Severino, S. A. Tennille, A. Moore, W. Jones, and J. Sanyal, "Smart mobility in the cloud: Enabling real-time situational awareness and cyber-physical control through a digital twin for traffic," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 3, pp. 3145–3156, 2023.

[18] Z. Wang, O. Zheng, L. Li, M. Abdel-Aty, C. Cruz-Neira, and Z. Islam, "Towards next generation of pedestrian and connected vehicle in-the-loop research: A digital twin co-simulation framework," *IEEE Transactions on Intelligent Vehicles*, vol. 8, no. 4, pp. 2674–2683, 2023.

[19] K. Duran and B. Canberk, "Digital twin enriched green topology discovery for next generation core networks," *IEEE Transactions on*

This article has been accepted for publication in IEEE Open Journal of Vehicular Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/OJVT.2024.3484956

Green Communications and Networking, vol. 7, no. 4, pp. 1946–1956, 2023.

[20] E. Ak, K. Duran, O. A. Dobre, T. Q. Duong, and B. Canberk, "T6conf: Digital twin networking framework for ipv6-enabled net-zero smart cities," *IEEE Communications Magazine*, vol. 61, no. 3, pp. 36–42, 2023.

[21] J. Yang, F. Lin, C. Chakraborty, K. Yu, Z. Guo, A.-T. Nguyen, and J. J. P. C. Rodrigues, "A parallel intelligence-driven resource scheduling scheme for digital twins-based intelligent vehicular systems," *IEEE Transactions on Intelligent Vehicles*, vol. 8, no. 4, pp. 2770–2785, 2023.

[22] R. C. Antúnez and M. W. Levin, "How does eco-routing affect total system emissions? city network predictions from user equilibrium models," *IEEE Transactions on Intelligent Transportation Systems*, vol. 25, no. 2, pp. 1408–1417, 2024.

[23] M. Singh and R. K. Dubey, "Deep learning model based co2 emissions prediction using vehicle telematics sensors data," *IEEE Transactions on Intelligent Vehicles*, vol. 8, no. 1, pp. 768–777, 2023.

[24] Y. Sun, Y. Hu, H. Zhang, F. Wang, and H. Chen, "A parallel supervision system for vehicle co2 emissions based on obd-independent information," *IEEE Transactions on Intelligent Vehicles*, vol. 8, no. 3, pp. 2077–2087, 2023.

[25] Y. Ge, H. Li, and A. Tuzhilin, "Route recommendations for intelligent transportation services," *IEEE Transactions on Knowledge and Data Engineering*, vol. 33, no. 3, pp. 1169–1182, 2021.

[26] W. Zhu and C. Zhu, "A prediction-based route guidance method toward intelligent and green transportation system," *IEEE Internet of Things Journal*, vol. 10, no. 14, pp. 12 763–12 776, 2023.

[27] D. Li and J. Lasenby, "Mitigating urban motorway congestion and emissions via active traffic management," *Research in Transportation Business & Management*, p. 100789, 2022.

[28] K. Kušić, R. Schumann, and E. Ivanjko, "A digital twin in transportation: Real-time synergy of traffic data streams and simulation for virtualizing motorway dynamics," *Advanced Engineering Informatics*, vol. 55, p. 101858, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1474034622003160

[29] H. Li, Y. Chen, K. Li, C. Wang, and B. Chen, "Transportation internet: A sustainable solution for intelligent transportation systems," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 12, pp. 15 818–15 829, 2023.

[30] Y. Gao, P. Si, K. Jin, T. Sun, and W. Wu, "Performance comparison of different deep reinforcement learning algorithms for task scheduling problem in blockchain-enabled internet of vehicles," *IEEE Transactions on Vehicular Technology*, vol. 72, no. 7, pp. 9322–9336, 2023.

[31] K. Duran, M. Broadbent, G. Yurdakul, and B. Canberk, "Digital twin-native ai-driven service architecture for industrial networks," in *2023 IEEE Globecom Workshops (GC Wkshps)*, 2023, pp. 1297–1302.

[32] T. Tang, R. Liu, C. Choudhury, A. Fonzone, and Y. Wang, "Predicting hourly boarding demand of bus passengers using imbalanced records from smart-cards: A deep learning approach," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 5, pp. 5105–5119, 2023.

[33] H. Yang, S. Pan, P. Zhang, L. Chen, D. Lian, and C. Zhang, "Binarized attributed network embedding," in *2018 IEEE International Conference on Data Mining (ICDM)*, 2018, pp. 1476–1481.

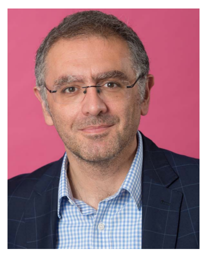
[34] L. Bellocchi, V. Latora, and N. Geroliminis, "Dynamical efficiency for multimodal time-varying transportation networks," *Scientific reports*, vol. 11, no. 1, pp. 1–14, 2021.

[35] C. GATELY, L. HUTYRA, and I. WING, "Darte annual on-road co2 emissions on a 1-km grid, conterminous usa v2, 1980-2017," 2019. [Online]. Available: https://daac.ornl.gov/cgi-bin/dsviewer.pl?ds$_i d = 1735$

[36] P. A. Lopez, M. Behrisch, L. Bieker-Walz, J. Erdmann, Y.-P. Flötteröd, R. Hilbrich, L. Lücken, J. Rummel, P. Wagner, and E. Wießner, "Microscopic traffic simulation using sumo," in *The 21st IEEE International Conference on Intelligent Transportation Systems*. IEEE, 2018. [Online]. Available: https://elib.dlr.de/124092/

KUBRA DURAN (Graduate Student Member, IEEE) received her B.Sc. degree in Electrical and Electronics Engineering in 2016 and her M.Sc. degree in 2021 in Computer Engineering from Istanbul Technical University. She has five years of experience in the industry, performing in R&D Specialist and Software Engineering roles. She has been awarded the Best MSc Thesis Award'22 by the IEEE Communications Society Turkey Chapter and the IEEE Computer Society Turkey Chapter. She is currently pursuing her Ph.D. in the School of Computing, Engineering and the Built Environment at Edinburgh Napier University. Her research interests are Digital Twin Networks and Intelligent Internet of Things.

LAL VERDA CAKIR (Graduate Student Member, IEEE) received her B.Sc. degree in Computer Engineering from Istanbul Technical University in 2022. She is currently pursuing her Ph.D. in the School of Computing, Engineering and the Built Environment, Edinburgh Napier University. Her research interest includes Digital Twins and AI-based Network Management.



ACHILLE FONZONE received the degree in civil engineering, the Ph.D. degree in transport and planning, and the master's Certificate in teaching and learning in higher education. He is currently a Professor of transport analysis and planning with Edinburgh Napier University. His current research interests include the area of travel attitudes and behavior in relation to smart and sustainable mobility. His research aims to promote a more equitable, cleaner, and safer mobility, making the most of the opportunities generated by automation, data and connectivity, and new means of transport and business models.

TRUNG Q. DUONG (Fellow, IEEE) is a Canada Excellence Research Chair (CERC) and a Full Professor at Memorial University of Newfoundland, Canada. He is also the adjunct Chair Professor in Telecommunications at Queen's University Belfast, UK and a Research Chair of Royal Academy of Engineering, UK. His current research interests include quantum communications, wireless communications, signal processing, machine learning, and realtime optimisation.

Dr. Duong has served as an Editor/Guest Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, IEEE TRANSACTIONS ON COMMUNICATIONS, IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, IEEE COMMUNICATIONS LETTERS, IEEE WIRELESS COMMUNICATIONS LETTERS, IEEE WIRELESS COMMUNICATIONS, IEEE COMMUNICATIONS MAGAZINES, and IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS. He received the Best Paper Award at the IEEE VTC-Spring 2013, IEEE ICC 2014, IEEE GLOBECOM 2016, 2019, 2022, IEEE DSP 2017, IWCMC 2019, 2023, and IEEE CAMAD 2023. He has received the two prestigious awards from the Royal Academy of Engineering (RAEng): RAEng Research Chair (2021-2025) and the RAEng Research Fellow (2015-2020). He is the recipient of the prestigious Newton Prize 2017.

BERK CANBERK(Senior Member, IEEE) is a Professor within the School of Computing, Engineering and The Built Environment at Edinburgh Napier University-UK, where he leads interdisciplinary research and initiatives in AI-enabled Digital Twins, IoT Communication, and Smart Wireless Networks. He's also an Affiliated Professor within the Department of Artificial Intelligence and Data Engineering at Istanbul Technical University (ITU) and Adjunct Faculty within the Department of Electrical and Computer Engineering at Northeastern University USA. He is also the Innovation Director of BTS Group, the biggest network automation and cloud computing company in Turkey. He is a distinguished recipient of the UK Royal Academy of Engineering's Global Talent Endorsement. He is an active Associate Editor at several world-leading academic journals such as IEEE Transactions on Vehicular Technology, Elsevier Computer Networks Journal, Elsevier Communication Networks Journal, and IEEE Communications Letters. He's actively involved in several conferences as TPC chair and Organizing Committee Member. He has been a Post-Doctoral researcher at Georgia Institute of Technology USA between 2011-2013. He received his PhD in Computer Science from ITU Turkey in 2011, his MSc in Telecommunications Engineering from the Chalmers University of Technology Sweden in 2005, and his BSc in Electrical Engineering from ITU in 2003. He was an Associate Professor at the Department of Computer Engineering at ITU between 2016-2021 and a full Professor between 2021-2022. He has been involved with several industrial research activities with leading technology companies worldwide, including research scholarship program funding with Google Deepmind, TUBITAK, BTS Group Turkey, Turkcell, Turkish Telekom, and Uniper Energy Germany.