
Software Practice and Experience on Smart Mobility Digital Twin in Transportation and Automotive Industry: Toward SDV-empowered Digital Twin through EV Edge-Cloud and AutoML

Jonggu Kang

School of AI Convergence, Sungshin Women's University, Seoul, Republic of Korea
E-mail: jonggu.kang@sungshin.ac.kr

Received 02 September 2024; Accepted 04 December 2024

Abstract

A digital twin is a virtual representation of a physical asset that serves as a pivotal convergence technology that facilitates real-time prediction, optimization, monitoring, control, and improved decision-making. It can be widely applied to various domains, such as automotive, manufacturing, logistics, and smart cities. The automotive industry, in particular, is actively integrating digital twins throughout the product life cycle, from research and development, production, sales, and services to enhance the overall customer experience. This paper presents insights and lessons learned on software practice and experience related to implementing smart mobility digital twins, focusing on the potential of transportation digital twins built from data collected by electric vehicles (EVs) with EV edge cloud and automated machine learning (AutoML). Despite current limitations in data sufficiency, we forecast that, as the SDV trend accelerates and the adoption of EVs increases, the digital

Journal of Web Engineering, Vol. 23.8, 1155–1180.

doi: 10.13052/jwe1540-9589.2385

© 2025 River Publishers

twin will become essential for the intelligent transportation system (ITS) in future smart cities, enabling accurate traffic predictions even in areas with limited road infrastructure. The successful integration of real-time data, high-performance prediction models, and automated service environments will enhance the effectiveness toward an SDV edge-empowered transportation digital twin.

Keywords: digital twin, internet of things, mobility, connected vehicle, software-defined vehicle, electric vehicle, smart city, transportation, automotive, edge cloud.

1 Introduction

The concept of a digital twin was first born in the aerospace domain when the US National Aeronautics and Space Administration (NASA) implemented a physical model of an early spacecraft as a simulation in a computing environment to predict and reflect problems that may occur during spaceflight [1, 2]. Since then, digital twins have been used primarily in advanced research technologies in the manufacturing industry and have begun to be used in many industries outside of manufacturing, including robotics, energy, healthcare and transportation domains. The scope of application is expanding from real-world objects to spaces, processes, people, and specific events [1, 3–5].

Recently, the digital twin has developed into a general term for all convergence technologies used to create a twin of a real object in the virtual world, analyze situations that may occur in advance, predict future possibilities, and control reality. In other words, it is developing into a convergence technology that allows (1) digital replicas of objects in reality, (2) performing analysis and simulation based on real-time or augmented data, and (3) diagnosing and optimizing the status of objects in reality, and predict, and solve problems. Industry interest and expectations for digital twin technology can be confirmed through research forecasts on the digital twin market [1, 6].

Digital twins for mobility in the transportation and automotive domains have been reported relatively rarely, where the industry is applying digital twin technology across the entire product life cycle stages, including research and development (R&D), production, sales, and services, to improve customer experience. Connected vehicles can also be a good data source for implementing and realizing mobility digital twins. However, previous

studies [3–5] have not consider the product development life cycle of vehicles.

In the recent transportation and automotive industry, the transition to software-defined vehicles (SDVs) and the expansion of eco-friendly electric vehicles (EVs) can be cited as the most important trends. SDVs have the advantage of being advantageous in autonomous driving and electrical systems and are likely to expand further based on EV platforms in the future [5, 7]. Thus, EVs can be seen as the most important trend in the industry. In addition, EVs are continuously increasing worldwide. Gartner predicts that they will account for more than 50% of automobile manufacturers' models by 2030¹.

In this paper, we explore the practice and experience of building a smart mobility digital twin based on data collected from EVs, which are currently on the rise and are expected to be important in the future, and have various strengths in digitalization such as autonomous driving, connectivity, and vehicle data collection. To this end, we describe a digital twin case that builds an EV edge cloud data platform environment that can continuously collect/analyze data generated from EVs as the Internet of Things (IoT) edges and perform predictive modeling using automated machine learning (AutoML).

Our contributions are as follows: (1) We describe the general aspects and actual implementation of digital twins in the transportation and automotive domain based on our experience. (2) We share insights and lessons learned from our implementation experiences throughout the product life cycle. (3) We evaluate the feasibility of collecting actual EV data and implementing services through the transportation digital twin case. (4) To the best of our knowledge, this is a practical case study of a smart mobility digital twin using real-time EV data with AutoML in the edge cloud environment for real application.

The remainder of this paper is organized as follows: Section 2 describes the background and related works. Section 3 presents software practice and experience on smart mobility digital twins. Section 4 explains a case study of smart mobility digital twins. Then, guidelines are presented through lessons learned based on digital twin projects in Section 5. Section 6 is a discussion of this study. Finally, this paper concludes in Section 7.

¹<http://www.gartner.com/en/industries/high-tech>

2 Background and Related Works

2.1 Edge Cloud

Edge cloud computing is a distributed computing system that integrates intelligence into smart mobility devices, allowing data to be processed and analyzed in real-time near the source of data collection. Thus, it improves response times for processing mobility big data through extensively distributed infrastructures. This low-latency responsiveness is recognized as important in an advancing modern society and is also important in the transportation and automotive domains [8–11]. Nevertheless, using in-vehicle controllers as edge intelligence is limited by safety and privacy issues, so, in this paper, we utilize devices that act as IoT edges attached to EVs and enable the data processing of EVs operating globally to be done close to the relevant region, thereby reducing latency and load.

2.2 Software-defined Vehicle (SDV)

The SDV is the up-to-date trend in smart mobility, where vehicle software is central to defining and updating various functions and performances. Software and hardware are decoupled, functions are enhanced through edge cloud, and continuous function improvement and maintenance are enabled through over-the-air (OTA) updates [7]. Figure 1 shows an example of the full architecture of an SDV.

2.3 Digital Twin

Many previous studies [1, 12–16] have extensively explored different perspectives on digital twins, such as definition, characteristics, applications, enabling technologies, and tools. Rasheed et al. [6] present the definition of a digital twin as follows: a virtual representation of a physical asset enabled through data and simulators for real-time prediction, optimization, monitoring, controlling, and improved decision making. They are widely applicable across diverse sectors, including automobiles, manufacturing, logistics, and smart cities [17–19]. Meanwhile, Kang [11] presented insights and lessons learned into software practices and experiences related to implementing digital twin projects. We extend Kang's study by presenting a transportation digital twin built based on data collected from EVs, which are currently on the rise and are expected to be important in the future, and have various strengths in digitalization, such as autonomous driving, connectivity, and vehicle data collection. Transportation digital twins can be used as an element of the

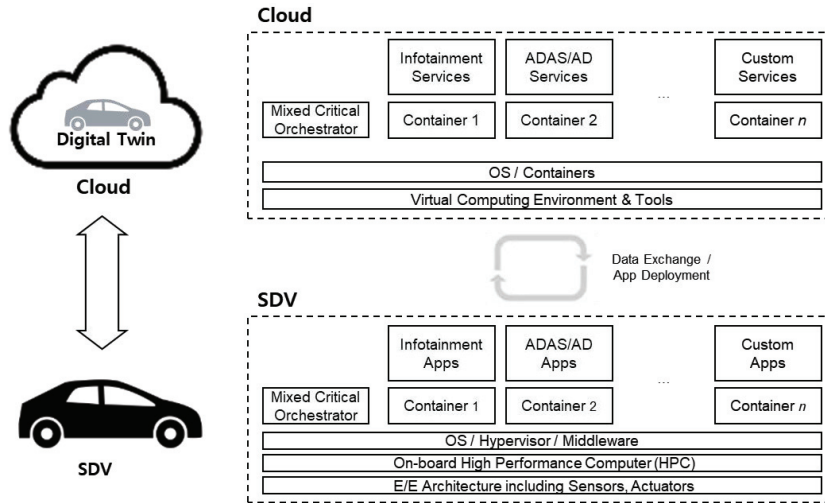


Figure 1 Full architecture of an SDV.

intelligent transportation system (ITS), which will play an important role in smart cities in the future, and will be used to predict traffic using only data collected from vehicles in countries with insufficient road infrastructure.

2.4 Related Works

Digital twins for mobility in the transportation and automotive domains have relatively rarely been reported compared to other domains such as manufacturing, aviation, healthcare, education, and cities [6, 12]. Jafari et al. [20] aimed to survey applications of digital twins in the development of the various aspects of energy management within a city, including power grids, microgrids, and transportation systems. The authors in [21] noted that the use of actual traffic data in real-time motorway analysis has not yet been explored and focused on simulation with real-time data integration during system runtime. Yan et al. [22] explored digital twin applications throughout the life cycle in the transportation engineering sector, segmented into design and optimization of infrastructure development, monitoring and management during the construction phase, intelligent infrastructure operation, and maintenance. Wu et al. [23] aimed to provide a reference for the subsequent intelligent development and construction of the transportation infrastructure in smart cities using digital twin and AI technology which showed significant advantages in the classification of transportation infrastructure and the management

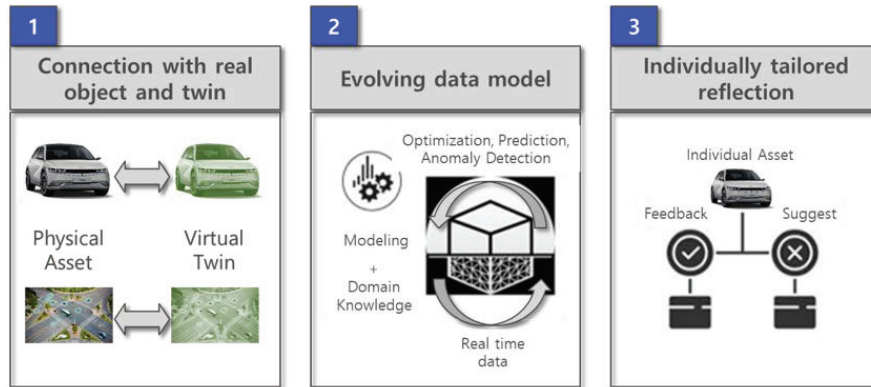


Figure 2 Major requirements of a smart mobility digital twin [11].

of a transportation spatial information network. Xu et al. [24] presented the design, implementation, and use cases of the digital twin toward the vision for smart city applications for supporting decision-making to reduce traffic congestion, incidents, and vehicle fuel consumption. Irfan et al. [25] focused on providing a comprehensive understanding of the requirements, reference architecture, challenges, and future research opportunities for a transportation digital twin system. Wang et al. [3] presented a mobility digital twin (MDT) framework. This MDT consists of three building blocks in the physical space (namely, human, vehicle, and traffic). A case is built with Amazon Web Services (AWS) to accommodate the proposed MDT framework.

Despite ongoing research efforts in the field, prior studies have focused on their own purpose, perspective, and enabling technologies. However, research on digital twins for smart mobility across the product life cycle has yet to be done. Since actual EV data are rare compared to traditional cars, there exists the possibility of discrepancies in the results obtained from previous studies. Consequently, our objective is to explore and evaluate the smart mobility digital twin using actual EV data with an EV edge cloud infrastructure and AutoML for real application.

3 Software Practice and Experience on a Smart Mobility Digital Twin

The starting point for building a digital twin is to specify requirements, as with typical software systems. In transportation and the automotive industry,

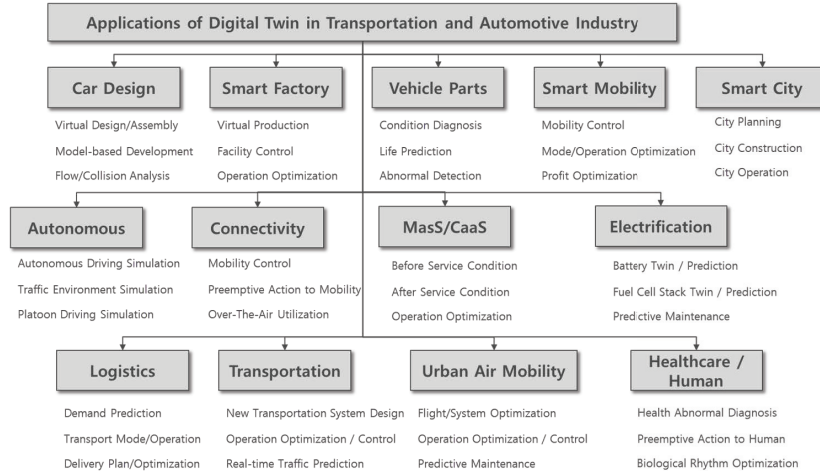


Figure 3 Applications of a digital twin in the transportation and automotive industry [11].

as shown in Figure 2, the general key functional requirements of digital twin are as follows:

- **Requirement 1** should create an identical twin in the virtual world by reflecting the characteristics of the physical object in the real world and connecting the object and the twin.
- **Requirement 2** should construct an infrastructure or a platform that can configure one or more evolving data models by reflecting continuously collected data in analysis and/or simulation.
- **Requirement 3** could provide a means to reflect the results derived from the model for each individual object (e.g., automatic feedback, recommendation, and suggestions for decision-making).

A digital twin that performs the above representative functional requirements has the following four characteristics:

- **Characteristic 1:** Two-way data flow where real-time data is stored/analyzed at the required rate, and the results are synchronized with reality.
- **Characteristic 2:** Seamless integration of data from connection with real physical objects for analysis, simulation, prediction, and derivation of solutions.
- **Characteristic 3:** Convergent use of contemporary element technologies such as IoT, cloud, big data, artificial intelligence, machine learning, augmented reality, virtual reality, etc.

- **Characteristic 4:** Can be applied broadly across business areas to make prediction-based decisions and optimization, as illustrated in Figure 3 [1, 3–5, 26, 27].

3.1 Reference Architecture

The reference architecture to realize the requirements is a combined framework of city [4], mobility [3], smart electric vehicles [5], and digital twin frameworks. The city framework proposed in [4], consists of three layers: (1) physical entities, (2) meta where digital twins and avatars are located with enabling technologies, and (3) applications layers. The MDT framework proposed in [3] consists of three components: (1) physical space including humans, vehicles, and traffic infrastructures, (2) digital space where the digital replicas of physical entities are activated, and (3) a communication plane to allow data flow for both directions. Three entities are considered in this MDT framework: Human, vehicle, and traffic. [5] presents a local and cloud-based architecture for smart electric vehicles. [25] also presented a reference architecture of transportation digital twin.

3.2 Steps and Enabling Technology

The digital twin starts with defining the requirements and ends with deriving the evolving results required by the final product service. It proceeds from the stages of data collection and transfer, data analytics, and service provision.

Each step illustrated in the upper part of Figure 4 is described as follows:

- **Step 1:** Define specific requirements: The process of defining the scope and purpose of the digital twin, selecting the target and scope of

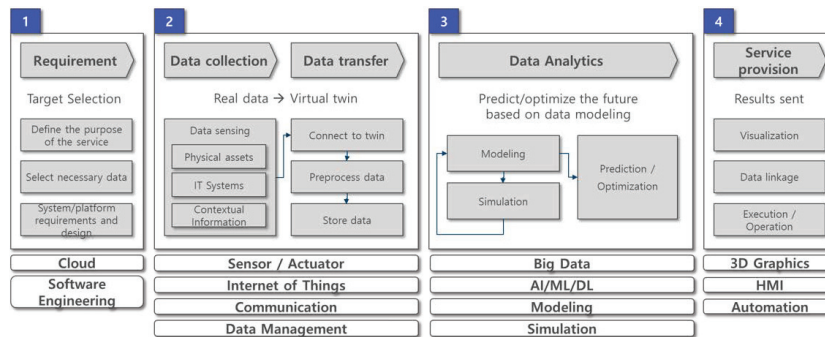


Figure 4 Steps and enabling technologies of a smart mobility digital twin [11].

implementation, and building a twin in the same form as the physical entity on the cloud as the realization of the virtual world.

- **Step 2:** Data collection/transfer: The process of collecting real-world data such as automobiles, human beings, buildings, factories, transportation infrastructure, and in-system information in real-time, and quickly and securely transmitting and storing them to the twin in the cloud using sensor/communication technologies.
 - Collect data in real-time and transfer it to the twin in the cloud. Real-world data is synchronized by connecting it with twins in the cloud.
 - Measure and collect data from sensors and actuators of physical objects and transmit the data through the process of communication, security, connection, pre-processing, and storage.
- **Step 3:** Data analytics: The process of analyzing data using various artificial intelligence, machine learning, and big data technologies as core of digital twin services and improving the accuracy of prediction models through repetitive simulation.
 - Modeling: A technique for finding and structuring complex real-world data features. It can be said that the core competencies are to structure all the characteristics related to objects, such as products, processes, people, spaces, behaviors, intentions and environments, integrate models, and find suitable models according to the unique characteristics of the data among numerous modeling methods, verify them, and improve their accuracy.
 - Simulation: Through tests on various conditions that are difficult to perform in the real world, it is possible to conduct analysis in the prediction and optimization stages. Simulation of models that reflect the real world can reduce error rates, uncertainties, and costs, and can include predictions using statistics, probability, artificial intelligence, and machine learning.
- **Step 4:** Service provision: The process of visualizing, linking, executing, and operating the results in accordance with the purpose of the service and as the final step to provide the results in the form of an actual service. It is a process that converts the analyzed results into information necessary for the product service and transmits them to support actual service decision-making, including visualization, data linkage, and execution/operation.

- Visualization: This is the process of converting and transmitting the analyzed results into information required for the service so that users can understand them by means of metaverse, web, or app.
- Data linkage: This is a service that links data to the final service, and it is possible to link with other systems, send APIs or messages, and combine with other platforms/services.
- Execution/operation: Manage service resources and perform operation and management for service maintenance such as inspection, evaluation, and fault detection.

As enabling technologies listed in the lower part of Figure 4 have been advanced, the digital twin will be realized up to “fully autonomous” without human intervention by recognizing and solving problems through complex data modeling and simulations [1, 26, 27].

3.3 Emerging Results

In this section, we will explain emerging results of the smart mobility digital twin illustrated in Figure 5, within a transportation and automotive company, for each product life cycle, by dividing them into three stages: R&D, production, and sales and service [3–5, 26, 27].

The emerging digital twin in R&D is utilized to respond to trends such as autonomous driving, connectivity and SDV, and build a preemptive verification platform for software logic and functions. Digital twins are used for the virtual design of prototypes and virtual verification in a virtual driving environment. Car makers reported cases of quality control of car design in a virtual space using a VR headset rather than producing a number of costly car prototypes. Virtual verification is used to design a new SDV or electric vehicle model, especially as the proportion of software or artificial intelligence-based modules increases according to SDV trends. It refers to a method of verifying modules by creating virtual controllers, components, and vehicles instead of using real ones, allowing for complex functions to be verified from the early stage of development. The verified software is continuously updated on the vehicle in operation through the OTA service. Vehicle data is monitored in real-time, and the accumulated big data can be used for function improvement or preventive evolution activities for residual defects.

The emerging digital twin in the production and manufacturing process is establishing a system of systems to virtually verify and provide feedback on process and logistics scenarios with the goal of a fully autonomous

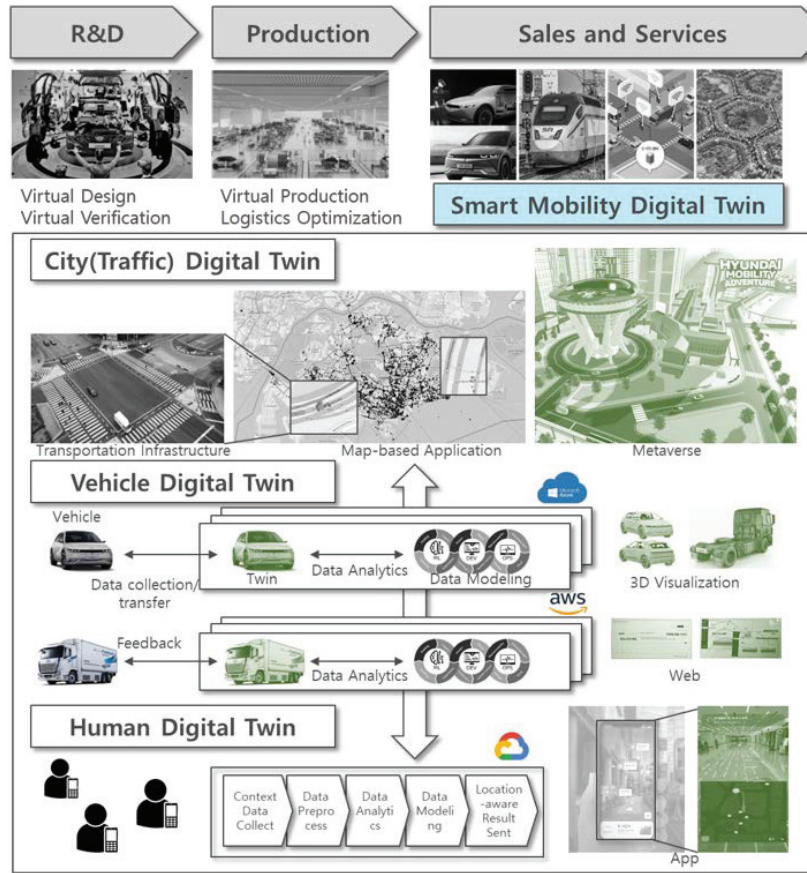


Figure 5 Emerging results of a smart mobility digital twin [11].

production system. Many car makers have reported on their smart factories, including digital twins. Digital twins can be used to design efficient processes by creating a virtual process identical to the actual process and repeating experiments on productivity changes. They can also be used to establish improvement directions. Utilizing spatial data, it is possible to reduce worker movement or lower production costs by designing an efficient workspace through simulation by replicating the production workspace.

The emerging result in sales and services is the smart mobility digital twin, used to increase customer satisfaction and service value in sales, use, management, and after-sales service processes. It is used to provide a proactive service by collecting real-time data from customers' vehicles, customers

themselves, and customers' surroundings, as well as providing analysis and prediction by means of metaverse, 3D, web, and app visualization. A mobility service provider twinizes smart electric vehicles and uses artificial intelligence, machine learning, and MLOps technology to comprehensively analyze actual driving data such as charging/discharging, driving habits, parking, and the environment that can affect a vehicle's performance. It is being used to support performance management, such as recommending customized vehicle and parts management plans for each vehicle. It also performs monitoring, regular inspection, and parts maintenance of railway/urban air mobility and passenger cars to provide safety and punctuality in virtual space and utilizes them to save time and cost. It also builds a next-generation ITS that delivers driving environment information to vehicles and supports safe driving through communication between road infrastructure and vehicles, and is used for autonomous driving testing, provision of road surface condition information, and road toll charging. To this end, the mobility service provider is gradually pursuing attempts to streamline customized services and operations for each target purpose by linking individual vehicle twins and spatial information and simulating traffic, logistics, and urban environments.

4 Smart Mobility Digital Twin using an EV Edge Cloud and AutoML

In the recent transportation and automotive industry, the transition to SDVs and EVs is regarded as the most promising trend. SDVs are likely to expand further based on EV platforms in the future due to the advantage of autonomous driving and electrical systems. In this chapter, we discuss the potential of a transportation digital twin built on data collected from EVs, which are currently on the rise and are expected to be important in the future, and have various strengths in digitalization, such as autonomous driving, connectivity, and vehicle data collection. To this end, we describe a case study based on a transportation digital twin that builds an EV edge cloud data platform environment that can continuously collect/analyze data generated from EVs as an IoT edge and perform AutoML predictive modeling by analyzing the collected data. Despite the fact that we also provide smart mobility services for battery and power-related predictions and driver-tailored guidance recommendations using the same edge-cloud infrastructure and the same EV data, this paper focuses on the transportation digital twin that performs traffic speed prediction.

4.1 EV Data Analytics

We first analyzed the characteristics of collected EV data for the purpose of using the data of EVs actually driven by drivers and predicting the speeds that are effective in understanding road traffic.

4.1.1 Data Description

We collected actual driving data directly by utilizing the infrastructure that can collect the necessary sensor data from 60 EVs. To this end, a terminal that can collect data from the network in the EV and transmit the data to the cloud was attached. Data for basic monitoring was collected at 10 s intervals, and data required for analysis was collected at 1 s intervals. The entire period was approximately 16 months, from November 2021 to April 2023, driven in the urban area of a large city. The features judged to be related to vehicle speed were preferentially selected from the entire EV data. Five engineers and data analysts with domain knowledge participated in this selection process. The selected features corresponded to the speed were approximately 300 points. The 300 selected points include accelerator pedal depth, brake light on status, wheel speed, battery voltage, current, temperature, etc. We note that the data was collected with the customer's prior consent, and the collected data was processed, stored, and analyzed without leaving the country to comply with local laws. We first conducted an analysis using only the data collected from EVs, and we used third party and external data to supplement the data shortage according to the results of the analysis. To supplement the data shortage and verify the collected data, TomTom data collected as real-time driving data of vehicles through IoT devices was processed and used as external big data.

4.1.2 Building an EV Edge Cloud Environment

As mentioned in the previous subsection, a total of 60 EVs were used in this study. Actual data generated from EVs are transmitted to a cloud environment through IoT edge devices. The cloud environment used Microsoft's Azure cloud environment because it is leading the large language model (LLM) through recent collaboration with OpenAI and is also leading in AI services with Databricks analysis tools.

Figure 6 shows a simplified structure of the EV edge cloud data collection and analysis environment built and used in this study. Although two-way communication is implemented, this paper only describes the architecture for EV data collection for security reasons. Each EV contains a

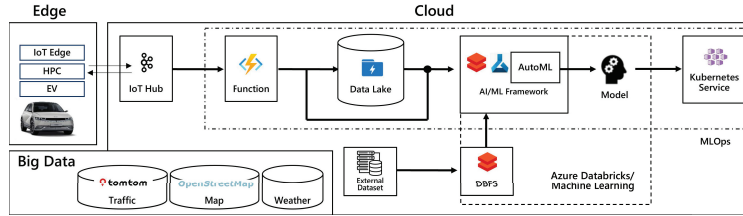


Figure 6 EV edge cloud data platform architecture.

high-performance computer (HPC), as it becomes an SDV, and an IoT edge. Data collected from the EV edge is stored in Azure Data Lake storage (ADLS) after basic pre-processing using Azure Function through the Kafka module in the cloud in the form of messaging. Data stored in ADLS can be analyzed and visualized using Azure Databricks and Azure machine learning analysis tools, and modeling for prediction can be performed. We utilized the AutoML services provided by Azure Databricks and Azure machine learning. Various services can utilize the generated prediction model in the form of an API via Azure Kubernetes service (AKS). Overall, the MLOps environment was built into the infrastructure so that model training and re-training can be continuously performed through automated data pipelines and training pipelines. In addition, in cases where analysis is difficult with only data collected from EVs, a pipeline and Databricks file system (DBFS) for external datasets was added to utilize third party big data to supplement the lack of data.

4.1.3 Performance of the AutoML Model

To obtain new insights from EV data, we utilized the AutoML service provided by Azure Databricks and Azure machine learning. The AutoML service is an automated machine learning service that automatically generates a model with the best performance by applying several pre-implemented machine learning algorithms to a given dataset, including a hyperparameter tuning process [28]. AutoML was run to predict the vehicle speed for EV data through Databricks and the Azure Machine Learning Studio. As a result, each unique ensemble model was recommended as the model with the best performance. The well-known root mean square error (RMSE) was used as a performance indicator.

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2}. \quad (1)$$

The best-performing model had an RMSE value of 23.19 in Databricks and an RMSE value of 19.13 in Azure ML. The performance of the model generated with AutoML achieved worse results compared to other traffic prediction studies [29, 30]. Thus, we confirmed that there was still room for improvement compared to the results of performing predictions with only the current EV data, so we decided to perform the analysis by adding external big data. We explored the minimum data size to reflect the actual road traffic environment. By synthesizing the opinions of domain engineers, data analysts, and other application information from the industry, we deduced that the number of sample sizes could reflect the actual road environment if it was greater than 10–17% of the actual [31]. The currently collected EV data was found to be the highest at 8% in the same time zone, and it was found that the smaller the value was, the greater the error between the actual speed and the EV speed difference. Therefore, we determined that the current dataset from 60 EVs was insufficient to represent the road environment. Therefore, we confirmed that accurate prediction is not viable with only the data collected from the current 60 EVs, so we decided to proceed with the analysis by adding external big data.

4.2 Considerations and Enhancement for a Transportation Digital Twin

In this section, we will examine the potential of a transportation digital twin built on data collected from EVs. In addition to using spatial information mainly used for transportation, this paper will describe the acquisition of real-time EV data as a source, a high-performance prediction model, and the environment for services.

4.2.1 EV data as a data source

We examine actual EV data from the perspective of data that enables the transportation digital twins. First, this study collected approximately 300 IoT sensor data of EVs based on the driving data of 60 EVs located in the urban area. Analysis was conducted using only EV data, as in Section 4.1, and it was found that there were shortcomings in prediction performance. To supplement this deficiency, TomTom data collected as real-time driving data of vehicles through IoT devices was used as external big data to supplement and verify the lack of data. We compared and analyzed EV data and TomTom data to see which set of data showed the best AutoML model performance when learning the baseline models. The combined method was as follows:

Table 1 Performance result for each combined dataset

	EV only	EV+TomTom (1)	EV+TomTom (2)	TomTom only
AutoML (1)	23.19	6.3847	8.7682	5.2619
AutoML (2)	19.13	6.7400	9.1128	6.4803

Table 2 Performance result for each combined model

	TFT	N-HITS	DeepAR	Stack Ensemble Model
RMSE	3.8036	4.8969	1.0569	1.4245

1. Dataset that consists only of EV data
2. Dataset that averages the speed of EV data and TomTom data
3. Dataset that joins the empty time of EV data with TomTom data
4. Dataset that consists only of Tomtom data.

As shown in Table 1, when testing the speed prediction model with the TomTom data set, the RMSE value was optimal, which can be judged as the best prediction. The dataset averaged with EV and Tomtom showed the next best performance. The EV-only dataset showed the worst performance, and the prediction was lower than that of other data sets. For now, the TomTom data set that was previously collected is judged to be the most appropriate for predicting road speed in the environment.

The amount of data collected from the 60 EVs used in this study was found to be insufficient for prediction. Based on the coverage study [31] on the estimated number of trips required to cover more than 80% of a city center through mobility AIoT services operating in large global cities, it is expected that the number of EV trips that can be collected will increase sufficiently as EVs increase in the near future while utilizing existing accumulated big data, sufficient data can be secured to build a transportation digital twin.

4.2.2 High-performance model as a prediction model

In this subsection, we investigate the possibility of improving the prediction performance through more sophisticated modeling. We predicted the average speed of a road where enough data was collected by selecting three well-known methods for time series prediction: RNN-based, MLP-based, and transformer-based models, respectively. The models we selected in this process are TFT [32], N-HITS [33, 34], and DeepAR [35]. This paper omits detailed descriptions of each model since they are beyond the scope of this paper. Research on traffic prediction models using open datasets can be found in [29].

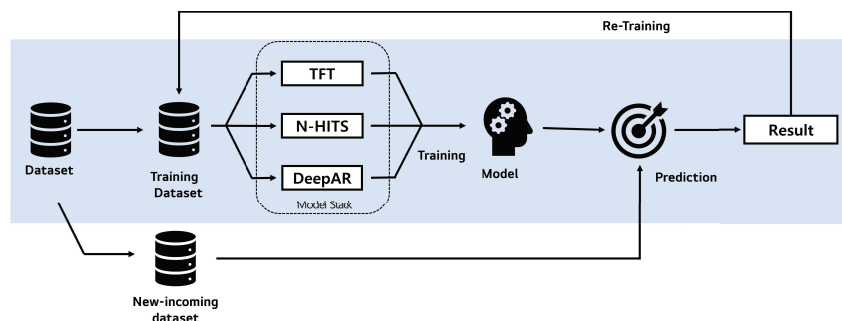


Figure 7 Stack ensemble model architecture.

In this study, after verifying the performance of each model as shown in Figure 7, we utilized an ensemble model that combines the strengths of the three individual models to improve the prediction performance. Each model predicted the average speed by the roadside (upbound: 1, downbound: 0) and the road segment (three road segments) in 15 min intervals. The final prediction value was generated by combining and averaging the results of each individual model. The reason for using an ensemble model that combines three models is that it can utilize the strengths of each model and complement its weaknesses. Model synthesis can prevent overfitting of individual models and provide generalized predictions. In addition, since it combines the predictions of multiple models, it can provide more diverse and suitable predictions than a single model. As a result, this method achieved 1.4245 RMSE, shown in Table 2, which confirmed the possibility of improving the prediction accuracy through the model.

Through these results, we were able to see that if we design the prediction model elaborately through the results of collecting and analyzing vehicle data, we can achieve sufficient prediction accuracy to build a transportation digital twin.

4.2.3 Automated ML as a service

Finally, we explore whether we could establish a system that automatically learns predictive models and deploys the models to perform services simultaneously with data collection.

In Figure 8, the speed of each section of the road was predicted using the stacking ensemble model, and the actual values were visualized for comparison. It was re-trained with new incoming data from 17 January to 30 January 2023 (2 weeks), and the average speed for every 15 min on

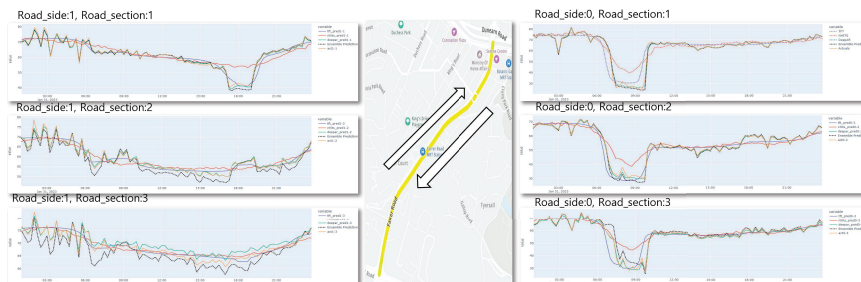


Figure 8 Results of speed prediction and visualization for each road section.

31 January was predicted. The prediction model showed that it accurately predicted a decrease in speed during the rush hour on the downbound lane from 7 to 10 AM. In addition, it was predicted that the speed of the upbound lane during rush hour from 17:00 to 19:00 would decrease as vehicles exited the road.

Overall, the MLOps environment was built into the infrastructure so that model training and re-training could be continuously performed through automated data pipelines and training pipelines. A prediction model was created on Azure Databricks, and the created model was deployed through AKS Service. First, a model was created and saved in Databricks. Afterward, the models saved in Databricks were loaded and registered in Azure Machine Learning Studio. After that, AKS was created, and the models registered in Azure ML were deployed through AKS. Kubernetes was used to consider future expansion in terms of scalability, high availability, and automation.

In terms of automated ML service, it was confirmed that MLOps service could be provided through the EV edge cloud environment, including automated data pipelines and training pipelines.

5 Lessons Learned

Based on the experience of digital twin projects in the transportation and automotive industry, the following guidelines are proposed for practitioners to conduct digital twin projects and apply them in their own business areas in the future.

- Using the edge cloud is a good option. Edge cloud is a good solution available at the time as it collects real-time or non-real-time data from objects, creates models with the data, builds XXOps such as dataOps, MLOps devOps, and secOps, and fuses multiple enabling

technologies into one platform, geographically located close to the user or data source. The edge cloud achieves all four characteristics of digital twins, including seamless integration and convergent use of contemporary element technologies. This study outlines deploying an edge-cloud architecture on Microsoft Azure but is not limited to Azure. The choice of the cloud platform is regarded as crucial as it must ensure the security of private data. Public cloud platforms, while convenient, come with inherent security and privacy risks due to the loss of control over resources. In contrast, private cloud platforms offer enhanced security and privacy by dedicating resources exclusively to a single organization, allowing for greater customization. However, private clouds are generally more expensive and less scalable than public clouds, which can offer on-demand resources and greater reliability through extensive server networks. The decision between public and private cloud platforms should be based on the specific purpose and needs of the digital twin services, with a combination of edge, hybrid, and multi-cloud approaches offering another viable option for building digital twins.

- A digital twin is not an all-purpose solution. Consider the cost. Costs should be considered in advance. Even though enabling technologies like cloud and artificial intelligence have become more common, more advanced, and cheaper, building a digital twin solution is still expensive. It is necessary to sufficiently review in advance whether a digital twin is the right choice. According to our internal survey, for one vehicle, the variable costs are still 100 times more expensive than expected relative to the goal of being mass-produced or attractive to customers.
- Decision-making and support from top management are essential. To realize a digital twin, multi-collaboration between business, R&D, production, service, ICT, strategy, and innovation organizations is essential. Therefore, top-down decision-making and support are essential. In addition, internal and external domain experts and industry–academia–research linkages may be necessary.
- Standardization of data, communication, and vocabulary is needed. Efforts on standardization are essential to accelerate multi-collaboration as well as internal and external domain experts, and industry–academia–research linkages. Standards are needed in the transportation domain to fully implement digital twin technology in connected vehicles. These standards would enable secure data management and interaction between different transportation entities and help in designing better

customer satisfaction services. However, achieving standardization in the transportation domain may face challenges due to differing interests between the transportation sector and the automotive sector.

- 3D modeling or visualization is not always necessary. Depending on the target and purpose of the digital twin, the combined technologies and final results are different.
- The focus is on model evolution and individualized feedback. Regardless of the target and purpose of the digital twin, this statement is usually true.
- Above all, connecting the value chain is the most difficult task. To build a digital twin across the entire value chain that connects R&D, production, and services, co-operation and collaboration between many organizations and stakeholders, and continuous promotion, are very important. This is only made possible when top management's decision-making and support, standardization, and convergence of enabling technologies are properly harmonized. Moreover, it takes a long time due to slow and step-by-step progress, so stable and patient support is needed.

6 Discussion

In this section, we would like to mention additional concerns.

- **Privacy.** There have been many concerns about personal information. This means that data collection and utilization must be done with the consent of the individual for the service. Security must be given special attention in data storage and utilization, and sensitive information must be safely utilized through encryption or anonymization.
- **Scalability.** This study implemented edge intelligence by attaching additional IoT devices to the vehicle. With the advent of the SDV and the use of HPC, it is expected that attempts to safely utilize additional computing in vehicles will be possible in the future. SDVs with true edge intelligence and edge cloud configurations are expected to be possible. Amid the recent global chaos caused by cloud failures, continuous service and operation utilizing edge-based delegation of authority and multi-cloud are expected to receive more attention.
- **Regulation.** Despite the above technical possibilities, computing may not be permitted when the vehicle is not in use or may be restricted by local regulations.

7 Conclusion

The paper discusses insights and guidelines on developing smart mobility digital twins, focusing on their application in transportation and automotive domains. The potential of fully autonomous digital twins, synchronized with big data such as urban spatial data and sensor information, is highlighted for advancing smart mobility and city services. However, more research is needed, especially as digital twins expand across different applications throughout the development life cycle.

The future will likely see an increasing demand for diverse digital twin applications, particularly in transportation, where they can enhance customer convenience and value through comprehensive analysis of road environments and mobility data. Despite the limited data from EVs in the current study, the growing prevalence of SDVs and EVs, which are well-suited for digitalization, offers significant potential for developing transportation digital twins. The rise of SDVs and EVs will further drive the collection of mobility data and utilization of edge intelligence.

In conclusion, for these digital twins to be effective, it's crucial to integrate real-time data, develop high-performance prediction models, and create automated service environments. They are expected to be key components of ITS in future smart cities, capable of performing traffic predictions even in areas with limited infrastructure. As a smart mobility digital twin in transportation and automotive domains evolves toward an SDV edge-empowered transportation digital twin, future studies are expected to address these real-world challenges in terms of cost, privacy, scalability, regulation, and standardization.

Acknowledgement

The contents of this paper only reflect the advanced and innovative views of the author, based on software practice and experience of digital twin projects. The contents do not necessarily reflect the official strategy and mass production views of Hyundai Motor Group. The author also wants to sincerely thank the former digital twin working group members, Jaehoon Sim, Joohyun Lee, and Jennie Kim, and Microsoft, Cloocus for their effort and support. This work was supported by the Sungshin Women's University Research Grant of 2024.

References

- [1] Aidan Fuller, Zhong Fan, Charles Day, and Chris Barlow. Digital twin: Enabling technologies, challenges and open research. *IEEE access*, 8:108952–108971, 2020.
- [2] Maulshree Singh, Evert Fuenmayor, Eoin P Hinchy, Yuansong Qiao, Niall Murray, and Declan Devine. Digital twin: Origin to future. *Applied System Innovation*, 4(2):36, 2021.
- [3] Ziran Wang, Rohit Gupta, Kyungtae Han, Haoxin Wang, Akila Ganlath, Nejb Ammar, and Prashant Tiwari. Mobility digital twin: Concept, architecture, case study, and future challenges. *IEEE Internet of Things Journal*, 9(18):17452–17467, 2022.
- [4] Ibrar Yaqoob, Khaled Salah, Raja Jayaraman, and Mohammed Omar. Metaverse applications in smart cities: Enabling technologies, opportunities, challenges, and future directions. *Internet of Things*, page 100884, 2023.
- [5] Ghanishtha Bhatti, Harshit Mohan, and R Raja Singh. Towards the future of smart electric vehicles: Digital twin technology. *Renewable and Sustainable Energy Reviews*, 141:110801, 2021.
- [6] Adil Rasheed, Omer San, and Trond Kvamsdal. Digital twin: Values, challenges and enablers from a modeling perspective. *IEEE access*, 8:21980–22012, 2020.
- [7] Zongwei Liu, Wang Zhang, and Fuquan Zhao. Impact, challenges and prospect of software-defined vehicles. *Automotive Innovation*, 5(2):180–194, 2022.
- [8] Florian Daniel and Federico Michele Facca. *Current Trends in Web Engineering, ICWE 2010 Workshops: 10th International Conference, ICWE 2010 Workshops, Vienna, Austria, July 5-6, 2010, Revised Selected Papers*, volume 6385. Springer Science & Business Media, 2010.
- [9] Giovanni Toffetti. Web engineering for cloud computing: (web engineering forecast: Cloudy with a chance of opportunities). In *Current Trends in Web Engineering: ICWE 2012 International Workshops: MDWE, ComposableWeb, WeRE, QWE, and Doctoral Consortium, Berlin, Germany, July 23-27, 2012, Revised Selected Papers 12*, pages 5–19. Springer, 2012.
- [10] Sundas Iftikhar, Sukhpal Singh Gill, Chenghao Song, Minxian Xu, Mohammad Sadegh Aslanpour, Adel N Toosi, Junhui Du, Huaming Wu,

- Shreya Ghosh, Deepraj Chowdhury, et al. Ai-based fog and edge computing: A systematic review, taxonomy and future directions. *Internet of Things*, 21:100674, 2023.
- [11] Jonggu Kang. Software practice and experience on smart mobility digital twin in transportation and automotive industry. In *the 4th International Workshop on Big data driven Edge Cloud Services (BECS 2024) Co-located with the 24th International Conference on Web Engineering (ICWE 2024), June 17, 2024, Tampere, Finland, Revised Selected Papers*. Springer, 2024.
- [12] Barbara Rita Barricelli, Elena Casiraghi, and Daniela Fogli. A survey on digital twin: Definitions, characteristics, applications, and design implications. *IEEE access*, 7:167653–167671, 2019.
- [13] David Jones, Chris Snider, Aydin Nassehi, Jason Yon, and Ben Hicks. Characterising the digital twin: A systematic literature review. *CIRP journal of manufacturing science and technology*, 29:36–52, 2020.
- [14] Mengnan Liu, Shuiliang Fang, Huiyue Dong, and Cunzhi Xu. Review of digital twin about concepts, technologies, and industrial applications. *Journal of manufacturing systems*, 58:346–361, 2021.
- [15] Fei Tao, Bin Xiao, Qinglin Qi, Jiangfeng Cheng, and Ping Ji. Digital twin modeling. *Journal of Manufacturing Systems*, 64:372–389, 2022.
- [16] Qinglin Qi, Fei Tao, Tianliang Hu, Nabil Anwer, Ang Liu, Yongli Wei, Lihui Wang, and Andrew YC Nee. Enabling technologies and tools for digital twin. *Journal of Manufacturing Systems*, 58:3–21, 2021.
- [17] Laisen Nie, Xiaojie Wang, Qinglin Zhao, Zhigang Shang, Li Feng, and Guojun Li. Digital twin for transportation big data: a reinforcement learning-based network traffic prediction approach. *IEEE Transactions on Intelligent Transportation Systems*, 25(1):896–906, 2023.
- [18] Evanthia Faliagka, Eleni Christopoulou, Dimitrios Ringas, Tanya Politi, Nikos Kostis, Dimitris Leonardos, Christos Tranoris, Christos P Antonopoulos, Spyros Denazis, and Nikolaos Voros. Trends in digital twin framework architectures for smart cities: A case study in smart mobility. *Sensors*, 24(5):1665, 2024.
- [19] Hailin Feng, Haibin Lv, and Zhihan Lv. Resilience towarded digital twins to improve the adaptability of transportation systems. *Transportation Research Part A: Policy and Practice*, 173:103686, 2023.
- [20] Mina Jafari, Abdollah Kavousi-Fard, Tao Chen, and Mazaher Karimi. A review on digital twin technology in smart grid, transportation system and smart city: Challenges and future. *IEEE Access*, 11:17471–17484, 2023.

- [21] Krešimir Kušić, René Schumann, and Edouard Ivanjko. A digital twin in transportation: Real-time synergy of traffic data streams and simulation for virtualizing motorway dynamics. *Advanced Engineering Informatics*, 55:101858, 2023.
- [22] Bin Yan, Fan Yang, Shi Qiu, Jin Wang, Benxin Cai, Sicheng Wang, Qasim Zaheer, Weidong Wang, Yongjun Chen, and Wenbo Hu. Digital twin in transportation infrastructure management: a systematic review. *Intelligent Transportation Infrastructure*, 2:liad024, 2023.
- [23] Jingyi Wu, Xiao Wang, Yukun Dang, and Zhihan Lv. Digital twins and artificial intelligence in transportation infrastructure: Classification, application, and future research directions. *Computers and Electrical Engineering*, 101:107983, 2022.
- [24] Haowen Xu, Andy Berres, Srikanth B Yoginath, Harry Sorensen, Phil J Nugent, Joseph Severino, Sarah A Tennille, Alex Moore, Wesley Jones, and Jibonananda 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*, 24(3):3145–3156, 2023.
- [25] Muhammad Sami Irfan, Sagar Dasgupta, and Mizanur Rahman. Towards transportation digital twin systems for traffic safety and mobility: A review. *IEEE Internet of Things Journal*, 2024.
- [26] CK Lo, Chun-Hsien Chen, and Ray Y Zhong. A review of digital twin in product design and development. *Advanced Engineering Informatics*, 48:101297, 2021.
- [27] Chiara Cimino, Elisa Negri, and Luca Fumagalli. Review of digital twin applications in manufacturing. *Computers in industry*, 113:103130, 2019.
- [28] Xin He, Kaiyong Zhao, and Xiaowen Chu. Automl: A survey of the state-of-the-art. *Knowledge-based systems*, 212:106622, 2021.
- [29] Jingyuan Wang, Jiawei Jiang, Wenjun Jiang, Chao Li, and Wayne Xin Zhao. Libcity: An open library for traffic prediction. In *Proceedings of the 29th international conference on advances in geographic information systems*, pages 145–148, 2021.
- [30] Ghulam Mustafa, Seong-je Cho, and Youngsup Hwang. Trafficnet: A hybrid cnn-fnn model for analysis of traffic accidents in seoul. *Journal of Computing Science and Engineering*, 17(4):182–194, 2023.

- [31] Kevin P O’Keeffe, Amin Anjomshoaa, Steven H Strogatz, Paolo Santi, and Carlo Ratti. Quantifying the sensing power of vehicle fleets. *Proceedings of the National Academy of Sciences*, 116(26):12752–12757, 2019.
- [32] Bryan Lim, Sercan Ö Arık, Nicolas Loeff, and Tomas Pfister. Temporal fusion transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting*, 37(4):1748–1764, 2021.
- [33] Cristian Challu, Kin G Olivares, Boris N Oreshkin, Federico Garza Ramirez, Max Mergenthaler Canseco, and Artur Dubrawski. Nhits: Neural hierarchical interpolation for time series forecasting. In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, pages 6989–6997, 2023.
- [34] Boris N Oreshkin, Dmitri Carпов, Nicolas Chapados, and Yoshua Bengio. N-beats: Neural basis expansion analysis for interpretable time series forecasting. *arXiv preprint arXiv:1905.10437*, 2019.
- [35] David Salinas, Valentin Flunkert, Jan Gasthaus, and Tim Januschowski. Deepar: Probabilistic forecasting with autoregressive recurrent networks. *International journal of forecasting*, 36(3):1181–1191, 2020.

Biography



Jonggu Kang received his B.S. degree and M.Sc. degree in computer science from Korea Advanced Institute of Science and Technology (KAIST) in 2006 and 2008, respectively. He received his Ph.D. degree in computer science and future vehicle from KAIST in 2022. He worked in future mobility industries for over 15 years from 2008 to 2023. He worked as a senior researcher at Hyundai Heavy Industries, where he was responsible for leading IoT platform

1180 *Jonggu Kang*

initiatives, and as a senior research engineer at Hyundai Motor Company, where he was responsible for leading digital twin initiatives. His research activities and interests are focused on SDVs, future mobility, smart cities, transportation, digital twins, and related AIoT/LLM applications. He is currently an assistant professor in school of AI convergence at Sungshin Women's University. He is a member of the IEEE.