LLM-Based Telemetry Repair and Fault Detection in V2X Networks with Digital Twin Guidance

Bishmita Hazarika*, Keshav Singh‡, Berk Canberk§, and Trung Q. Duong*,†

*Memorial University of Newfoundland, St. Johns, Canada, e-mail:{bhazarika, tduong}@mun.ca

‡National Sun Yat-sen University, Kaohsiung 80424, Taiwan, e-mail: ksingh1980@ieee.org

§Edinburgh Napier University, Edinburgh EH10 5DT, UK, e-mail: b.canberk@napier.ac.uk

†Queen's University Belfast, UK, e-mail: trung.q.duong@qub.ac.uk

***Abstract*—In vehicle-to-everything (V2X) networks, real-time telemetry is essential for enabling predictive analytics and fault detection in intelligent transportation systems. However, frequent wireless disruptions due to interference, mobility, and congestion lead to telemetry gaps that degrade downstream decision-making. To address this challenge, we propose a framework that enhances wireless telemetry robustness using large language models (LLMs) guided by digital twin-based context. Our system combines retrieval-augmented generation with environmental priors to recover high-dimensional, time-correlated telemetry streams lost during communication outages. We also integrate federated continual learning to maintain fault classification performance across non-i.i.d. V2X conditions without centralized data exchange. Extensive evaluations on real-world driving datasets with simulated wireless impairments show that our method significantly improves reconstruction fidelity, reduces degradation from multi-step gaps, and sustains long-term classifier stability. This work demonstrates how AI-driven semantic recovery mechanisms can improve the functional reliability of wireless V2X telemetry under dynamic and lossy network conditions.**

I. Introduction

INTELLIGENT transportation systems increasingly rely on real-time telemetry from connected vehicles to support predictive analytics, anomaly detection, and safety-critical decision-making. In vehicle-to-everything (V2X) networks, this telemetry typically includes GPS trajectories, velocity, acceleration, and signal quality metrics that are periodically transmitted from vehicles to roadside units (RSUs). However, such networks are frequently subject to packet losses, wireless interference, and sensor outages that result in intermittent or missing telemetry streams. These disruptions significantly degrade downstream tasks such as fault classification, mobility prediction, and safety assurance. Traditional methods such as linear interpolation or recurrent imputation techniques are ill-suited for high-dimensional, temporally correlated, and context-sensitive telemetry, especially when the missing patterns are non-random or span multiple steps.

Recent advances in large language models (LLMs) have opened new possibilities for structured sequence modeling and intelligent telemetry recovery. In particular, large language models (LLMs), trained on structured sequences, have demonstrated impressive capabilities in autoregressive forecasting, multi-modal reasoning, and data imputation tasks [1], [2]. However, applying LLMs to spatio-temporal vehicular telemetry presents unique challenges. These include context sparsity, long-range dependencies, domain-specific dynamics, and the need for structured input formats. To overcome these limitations, Retrieval-augmented generation (RAG) frameworks [3] have been introduced, which guide generative models using external memory and contextual examples. While RAG-based prompting improves coherence and factual accuracy in natural language processing (NLP), its utility in real-time telemetry reconstruction remains underexplored.

Complementing generative capabilities with contextual awareness, digital twin (DT) technology offers a structured means to model and forecast the evolving physical state of a system. A DT aggregates spatial and temporal statistics such as vehicle density, velocity profiles, and signal reliability to summarize real-time environmental dynamics [4]. Prior work has demonstrated the use of DTs in industrial automation and predictive maintenance [5], but few studies have examined their integration into V2X telemetry repair pipelines. Notably, the combination of DT-derived priors with learned retrieval metrics offers a promising direction for conditioning LLMs on spatial-temporal consistency. On the classification side, federated learning (FL) [6] enables decentralized RSUs to collaboratively train fault classifiers without sharing raw data. However, conventional FL methods suffer from catastrophic forgetting and non-i.i.d. distribution shifts in vehicular environments. Federated continual learning (FCL) [7] extends FL by enabling incremental task learning and model stability over time, which is crucial for environments with recurring faults, evolving traffic patterns, and dynamic road conditions.

Despite these parallel advances, existing research has not unified LLMs, RAG, DTs, and FCL into a coherent architecture for resilient V2X telemetry recovery and fault classification. Prior works have either relied on sequence models without environmental grounding [8], used DTs without generative reconstruction [9], or performed federated classification without continual adaptation [10]. No existing framework jointly reasons over retrieval-enhanced generation, environment-conditioned recovery, and federated continual diagnosis under telemetry loss.

Motivated by the need for robust telemetry recovery and adaptive fault detection in lossy V2X settings, we propose

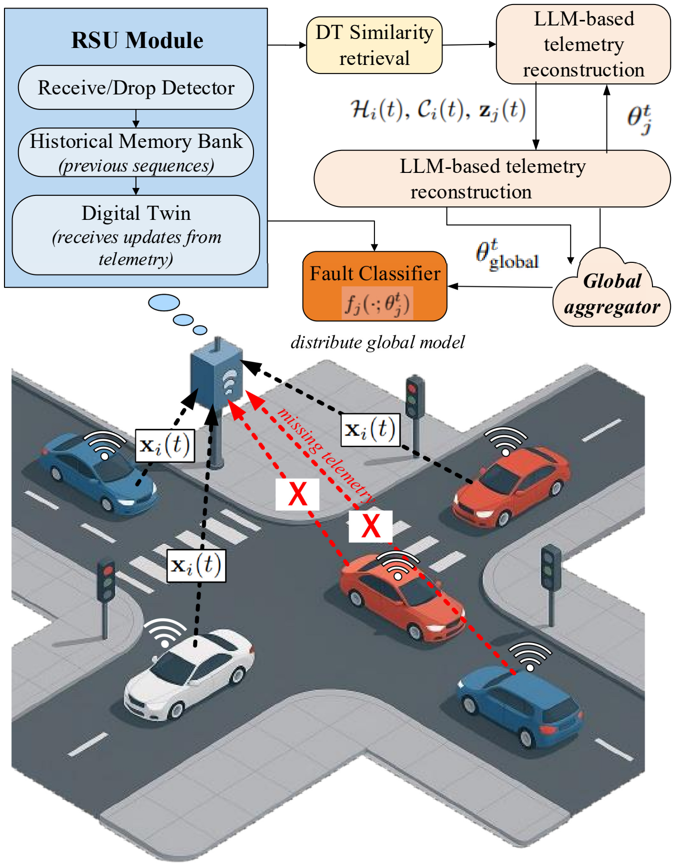

Fig. 1: An illustration of the proposed V2LLM framework.

V2LLM, a unified framework that integrates digital twin-driven retrieval, retrieval-augmented LLM prompting, and federated continual learning. V2LLM is designed to reconstruct missing telemetry vectors with high fidelity and to maintain fault classification accuracy across evolving driving contexts. The key contributions are:

- We develop a hybrid retrieval mechanism that combines trajectory similarity and digital twin-based environmental context to identify relevant past sequences from RSU memory, enabling context-aware prompt construction for LLM inference.
- We implement a retrieval-augmented LLM module that reconstructs multi-step telemetry gaps using autoregressive generation over structured prompts grounded in prior context and DT state.
- We integrate a federated continual learning classifier at each RSU to support non-i.i.d. telemetry and dynamic road conditions, preserving knowledge over sequential tasks via regularized updates.
- Through extensive simulation on preprocessed real-world telemetry (NGSIM and HighD), we demonstrate that V2LLM outperforms baselines in recovery fidelity, robustness under missing rates, drift resilience, and fault classification accuracy.

II. System Model

We consider a vehicular network composed of N vehicles $\mathcal{V} = \{v_1, v_2, \dots, v_N\}$ and R roadside units (RSUs) $\mathcal{R} = \{r_1, r_2, \dots, r_R\}$ deployed across a geographic region. Each vehicle $v_i \in \mathcal{V}$ is equipped with an onboard unit (OBU) that periodically transmits telemetry data to its nearest RSU via a vehicle-to-infrastructure (V2I) wireless link. A central cloud server aggregates model updates from all RSUs in a federated learning setup. Fig. 1 illustrates a small example of the considered V2X framework.

A. Vehicular Telemetry and Message Loss

Let T be the time horizon of the vehicular data collection process. At each discrete time $t \in \{1, 2, \dots, T\}$, a vehicle v_i generates a telemetry vector $\mathbf{x}_i(t) \in \mathbb{R}^d$, which includes mobility and communication metrics such that $\mathbf{x}_i(t) = [p_x(t), p_y(t), v(t), a(t), \theta(t), \text{RSSI}(t), \text{PLR}(t), \dots]$, where p_x, p_y are GPS coordinates, v is velocity, a is acceleration, and θ is heading. RSSI (received signal strength indicator) and PLR (packet loss rate) are communication-related metrics. Here, d denotes the total number of telemetry features. Let $\delta_i(t) \in \{0, 1\}$ be a reception indicator:

$$\delta_i(t) = \begin{cases} 1, & \text{if } \mathbf{x}_i(t) \text{ is received at RSU}, \\ 0, & \text{otherwise}. \end{cases} \tag{1}$$

Each RSU maintains a received buffer $\mathcal{X}_j(t) = \{\mathbf{x}_i(t) \mid \delta_i(t) = 1\}$ and a sliding historical window per vehicle given as $\mathcal{H}_i(t) = \{\mathbf{x}_i(t-k), \dots, \mathbf{x}_i(t-1)\}$. Each RSU is associated with a DT instance that maintains a continuously updated virtual replica of the local vehicular environment. The DT stores a high-fidelity representation of traffic dynamics, mobility patterns, and recent vehicular interactions within its coverage area. The integration of DT with is detailed in Section IV.

A data loss event is triggered when τ consecutive telemetry vectors from vehicle v_i are not received by time t. Formally, the RSU observes $\delta_i(t-\tau+1) = \cdots = \delta_i(t) = 0$. This defines a missing subsequence $\mathbf{x}_i(t-\tau+1 : t)$ which must be reconstructed. The recovery objective is to generate an approximation $\hat{\mathbf{x}}_i(t) \approx \mathbf{x}_i(t)$ using the historical context and relevant retrieved data.

B. Historical Knowledge Base for Retrieval

Each RSU r_j maintains a historical memory $\mathcal{K}_j$ consisting of M samples of valid past telemetry windows and their associated next-step ground truths:

$$\mathcal{K}_j = \left\{ \left(\mathcal{H}_i^{(m)}, \mathbf{x}_i^{(m)} \right) \right\}_{m=1}^{M}. \tag{2}$$

We define a weighting parameter $\alpha \in [0, 1]$ to balance trajectory-based similarity (e.g., velocity and position) and environment-based similarity (e.g., vehicle density, signal conditions). Given a new incomplete sequence $\mathcal{H}_i(t)$, a similarity function $s(\cdot, \cdot)$ (e.g., cosine similarity) retrieves the top-K most similar historical contexts:

$$\mathcal{C}_i(t) = \mathcal{R}(\mathcal{H}_i(t), \mathcal{K}_j) \subset \mathcal{K}_j. \tag{3}$$

C. LLM-Based Telemetry Recovery

The core of the V2LLM framework is a LLM that operates in a RAG mode to recover missing vehicular telemetry. When a data loss is detected, the RSU constructs a serialized textual prompt $\mathcal{Q}_i(t)$ based on: (i) the recent historical sequence $\mathcal{H}_i(t)$, (ii) The retrieved context set $\mathcal{C}_i(t)$, (iii) Metadata such as vehicle ID and timestamps.

The prompt $\mathcal{Q}_i(t)$ is constructed by flattening the historical telemetry vectors into a textual format. Each line corresponds to a time step and includes the feature vector. For example:

```
Vehicle ID: v_3
Time window: [t-3, t-2, t-1]
[t-3]: [52.1, 13.4, 11.2, 0.3, ..., -74.5, 0.02]
[t-2]: [52.3, 13.5, 11.4, 0.2, ..., -75.1, 0.01]
[t-1]: [52.6, 13.8, 11.7, 0.1, ..., --, --]
Predict next state:
```

Missing values are masked using a special token (e.g., ”–”) and inferred by the model. The prompt also includes K retrieved sequences from the RSU's memory bank.

To retrieve relevant context from the RSU's historical memory $\mathcal{K}_j$, the system computes the similarity between the current history window $\mathcal{H}_i(t)$ and each stored window $\mathcal{H}_i^{(m)}$:

$$s(\mathcal{H}_i(t), \mathcal{H}_i^{(m)}) = \frac{\langle \mathbf{h}_i, \mathbf{h}_i^{(m)} \rangle}{\|\mathbf{h}_i\| \cdot \|\mathbf{h}_i^{(m)}\|}, \tag{4}$$

where $\mathbf{h}_i$ and $\mathbf{h}_i^{(m)}$ are flattened versions of the input and candidate sequence windows. The top-K matching sequences form the retrieval set $\mathcal{C}_i(t)$.

The complete prompt $\mathcal{Q}_i(t)$, composed of historical context and retrieved samples, is fed into the LLM. The model then generates the missing telemetry vector as $\hat{\mathbf{x}}_i(t) = \mathrm{LLM}(\mathcal{Q}_i(t)) \in \mathbb{R}^d$. For consecutive missing steps $(t, t+1, \ldots, t+n)$, the LLM is applied in an autoregressive manner. That is, the generated output at each step is inserted into the prompt to assist the generation of the next step.

III. Problem Formulation

In highly dynamic vehicular networks (V2X), reliable data-driven fault diagnosis relies on complete, real-time telemetry data streams transmitted from vehicles to roadside infrastructure. However, such environments are prone to frequent packet loss, link degradation, or sensor malfunction, leading to intermittent telemetry gap. These disruptions hinder accurate fault classification and degrade the performance of downstream learning models.

Let $\mathbf{x}_i(t) \in \mathbb{R}^d$ denote the telemetry vector of vehicle v_i at time t, and let $\delta_i(t) \in \{0, 1\}$ be the indicator of successful reception at RSU r_j. When a contiguous window of telemetry vectors is missing, denoted by $\mathcal{M}_i(t) = \{\mathbf{x}_i(t'), \ldots, \mathbf{x}_i(t)\}, \quad \delta_i(t') = \cdots = \delta_i(t) = 0$.

In this study, our main objective is to reconstruct the telemetry gap $\mathcal{M}_i(t)$ with high fidelity such that downstream classifiers can operate as if the sequence were never lost.

Let $\mathcal{H}_i(t)$ denote the available telemetry history prior to loss, and $\mathcal{C}_i(t)$ be the retrieved context from historical data $\mathcal{K}_j$ at RSU r_j. We formulate the LLM-based recovery function as:

$$\mathcal{Q}_i(t) = \mathrm{Prompt}(\mathcal{H}_i(t), \mathcal{C}_i(t), \Phi_j(t)), \tag{5}$$

$$\hat{\mathcal{M}}_i(t) = \mathrm{LLM}(\mathcal{Q}_i(t)), \tag{6}$$

where the $\mathrm{Prompt}(\cdot)$ function constructs a structured text representation by flattening the telemetry history $\mathcal{H}_i(t)$, the retrieved context $\mathcal{C}_i(t)$, and the environmental summary $\Phi_j(t)$ into a token sequence compatible with the input requirements of the LLM. The generation objective is to minimize the reconstruction loss:

$$\mathcal{L}_{\text{recon}} = \sum_{t' \in \mathcal{M}_i(t)} \|\hat{\mathbf{x}}_i(t') - \mathbf{x}_i(t')\|_2^2 . \tag{7}$$

Furthermore, the DT associated with each RSU provides additional priors about traffic flow, vehicular density, and expected behavior under similar operational conditions. Let $\Phi_j(t)$ denote the DT-derived state summary at RSU r_j. This summary can be embedded into the LLM prompt or used to modulate retrieval scoring:

$$s'(\mathcal{H}_i(t), \mathcal{H}_i^{(m)}; \Phi_j(t)) = \lambda \cdot s(\cdot, \cdot) + (1-\lambda) \cdot \mathrm{sim}_{\text{DT}}(\Phi_j(t), \Phi^{(m)}), \tag{8}$$

where $\Phi^{(m)}$ is the historical DT state for the m-th window and $\lambda \in [0, 1]$ balances similarity sources. The final task is to perform accurate fault classification using both reconstructed and real telemetry:

$$\hat{y}_i(t) = f_j(\hat{\mathbf{x}}_i(t); \theta_j), \quad \hat{y}_i(t) \in \mathcal{Y}, \tag{9}$$

where f_j is the classifier at RSU r_j and $\mathcal{Y}$ is the set of fault types (e.g., normal, delay, spoofing, jamming). In the next section, we propose our integrated framework *V2LLM* that unifies digital-twin-aware retrieval, LLM-based generation, and federated continual learning for resilient V2X fault diagnosis.

IV. Proposed V2LLM Framework

In this section, we introduce the proposed *V2LLM* framework for LLM-driven DT and traffic-aware fault diagnosis in V2X networks. This framework integrates five interconnected layers: (1) vehicular telemetry and communication modeling, (2) DT-based state estimation, (3) RAG architecture for LLM prompting, (4) sliding-window autoregressive generation of telemetry vectors via LLM, and (5) FCL for distributed classification.

As defined in Section II, vehicles transmit telemetry $\mathbf{x}_i(t)$ to RSUs, with losses indicated by $\delta_i(t)$. Let $\mathcal{M}_i(t)$ denote a block of missing vectors:

$$\mathcal{M}_i(t) = \{\mathbf{x}_i(t'), \ldots, \mathbf{x}_i(t)\}, \quad \text{with } \delta_i(t') = \cdots = \delta_i(t) = 0. \tag{10}$$

The goal is to recover $\hat{\mathcal{M}}_i(t)$ using a hybrid context of real history and digital-twin-enhanced environment representation. The following subsections walk through the five key layers of V2LLM in their natural execution order:

A. Digital Twin-Driven State Estimation

Each RSU r_j maintains a DT model $\Phi_j(t)$ to capture real-time traffic and communication dynamics within its coverage. The DT is defined as:

$$\Phi_j(t) = (\mu_j(t), \sigma_j(t), \rho_j(t), \psi_j(t)), \tag{11}$$

where $\mu_j(t)$ is the moving average of vehicle speeds, $\sigma_j(t)$ is the estimated vehicle density, $\rho_j(t)$ denotes link reliability (from RSSI and PLR), and $\psi_j(t)$ is the lane occupancy vector derived from GPS reports. These components are updated as new telemetry arrives, and encoded into $\mathbf{z}_j(t) \in \mathbb{R}^{d_z}$, a compact representation of the local RSU state. This embedding provides environmental context to the LLM and enables consistency filtering for reconstructed telemetry.

B. RAG-based Prompt Construction

Once the environment is modeled via the DT, the RSU leverages it alongside historical telemetry to build context-rich prompts for LLM-based reconstruction. Each RSU r_j maintains a historical memory bank $\mathcal{K}_j$ containing recent valid telemetry windows and their corresponding next-step ground truth:

$$\mathcal{K}_j = \{(\mathcal{H}_i^{(m)}, \mathbf{x}_i^{(m)}, \Phi_j^{(m)})\}_{m=1}^{M}. \tag{12}$$

Given a telemetry gap for vehicle v_i at time t, the RSU constructs a query vector based on current history $\mathcal{H}_i(t)$ and digital twin state $\Phi_j(t)$. To retrieve similar sequences, we compute the joint similarity score between query and memory samples:

$$\begin{aligned}\text{sim}((\mathcal{H}_i, \Phi_j), (\mathcal{H}_i^{(m)}, \Phi_j^{(m)})) &= \alpha \cdot \cos(\mathcal{H}_i, \mathcal{H}_i^{(m)}) \\ &\quad + (1-\alpha) \cdot \cos(\mathbf{z}_j, \mathbf{z}_j^{(m)})\end{aligned} \tag{13}$$

where, $\alpha \in [0, 1]$ controls the weighting between trajectory similarity and environmental similarity. The top-K closest matches form the retrieved context set $\mathcal{C}_i(t)$. The prompt $\mathcal{Q}_i(t)$ for LLM inference is built by concatenating historical telemetry $\mathcal{H}_i(t)$, the retrieved contexts $\mathcal{C}_i(t)$, encoded DT context $\mathbf{z}_j(t)$, and metadata such as vehicle ID and timestamps.

The entire prompt is serialized into a structured format (e.g., JSON-like template or tabular format), and then tokenized for autoregressive LLM input. The use of RAG constrains generation to past-relevant behaviors, reducing hallucinations and ensuring alignment with the prevailing traffic context.

C. Autoregressive LLM Generation

Using the constructed prompt, the LLM autoregressively reconstructs missing telemetry values. The LLM receives the constructed prompt $\mathcal{Q}_i(t)$, a serialized mixture of structured telemetry, retrieved history, DT embeddings, and metadata, and generates the next-step telemetry vector via:

$$\hat{\mathbf{x}}_i(t) = \text{LLM}(\mathcal{Q}_i(t)) \in \mathbb{R}^d. \tag{14}$$

The prompt is formatted using a structured layout (e.g., a flattened JSON or CSV-like token sequence) and tokenized into subwords for autoregressive inference. The LLM may operate in zero-shot or few-shot settings, depending on fine-tuning availability. For multi-step gaps $[t, t+1, \dots, t+n]$, a sliding autoregressive generation is applied:

$$\mathcal{Q}_i(t+\ell) \leftarrow \mathcal{Q}_i(t+\ell-1) \cup \hat{\mathbf{x}}_i(t+\ell-1), \quad \ell = 1, \dots, n. \tag{15}$$

Each generated output is appended to the input prompt to guide the generation of the next time step. This continues until all missing vectors in $\mathcal{M}_i(t)$ are reconstructed. The decoder uses greedy decoding or sampling with temperature depending on deployment constraints, and output vectors are post-processed to ensure alignment with DT constraints (e.g., velocity bounds or heading continuity).

D. Federated Continual Learning

The reconstructed or received telemetry vectors are then used to train local classifiers at RSUs. To adapt to evolving road conditions and distributed data, we integrate FCL which is a distributed machine learning paradigm that combines the privacy-preserving benefits of federated learning with the adaptability of continual learning (CL). In the context of V2X, where telemetry data and operational environments evolve over time, traditional FL methods suffer from catastrophic forgetting and fail to retain knowledge from previous fault patterns or road scenarios. FCL addresses this by enabling each RSU to learn a sequence of non-i.i.d. tasks over time while contributing to a global model.

Let $\mathcal{T}_j^1, \mathcal{T}_j^2, \dots$ denote a task sequence at RSU r_j, where each task represents a new driving context or telemetry profile (e.g., day/night, urban/highway, normal/faulty behavior). Each RSU maintains a local classifier $f_j(\cdot; \theta_j^t) : \mathbb{R}^d \to \mathcal{Y}$, where $\mathcal{Y}$ is the label space of traffic conditions or fault classes. At each timestep t, the local model is updated using the following objective:

$$\min_{\theta_j^t} \underbrace{\mathcal{L}_j^t}_{\text{local loss}} + \lambda_1 \underbrace{\mathcal{R}(\theta_j^t)}_{\text{regularization}} + \lambda_2 \underbrace{\sum_{k=1}^{t-1} \left\| \theta_j^t - \theta_j^k \right\|^2}_{\text{stability to past tasks}}, \tag{16}$$

where $\mathcal{L}_j^t$ is the supervised loss on local data (either real or LLM-reconstructed), $\mathcal{R}$ is a task-specific regularization term, and the final term penalizes deviations from past parameters to preserve historical knowledge and promote temporal model stability. This regularization is inspired by elastic weight consolidation (EWC) and knowledge distillation methods such as learning without forgetting (LwF).

After a fixed number of local updates, RSUs transmit their current model parameters θ_j^t to the cloud, where a federated averaging mechanism is used to compute the global model:

$$\theta_{\text{global}}^t = \frac{1}{R} \sum_{j=1}^{R} \theta_j^t. \tag{17}$$

This aggregated model is then redistributed to all RSUs for continual learning in the next round. Compared to standard FL, this approach allows each RSU to specialize for its task

Algorithm 1 V2LLM Inference and Federated Learning Loop

```
Initialize: Local model θ_j^0 at each RSU r_j, DT Φ_j(t), memory K_j, global model θ_global^0.
for each time t and each vehicle v_i do
    if δ_i(t) = 0 then                                    ▷ Telemetry missing
        Retrieve H_i(t) and compute current Φ_j(t).
        Compute similarity via Eq. (13) to obtain C_i(t).
        Construct prompt Q_i(t) using H_i(t), C_i(t), z_j(t).
        Generate x̂_i(t) via Eq. (14).
        Update K_j with (H_i(t), x̂_i(t), Φ_j(t)).
    else
        Store x_i(t) in memory K_j.
    end if
    Feed x_i(t) or x̂_i(t) to classifier f_j(·; θ_j^t) to obtain prediction.
    Compute L_j^t and update θ_j^t via Eq. (16).
end for
After every T_sync rounds:
    Upload θ_j^t from each RSU to server.
    Aggregate global model via Eq. (17).
    Broadcast θ_global^t to all RSUs for next training cycle.
Output: Reconstructed telemetry x̂_i(t) and fault classification prediction for each v_i.
```

sequence while collaboratively constructing a generalizable global model across the V2X network. **Algorithm 1** describes the V2LLM interference and FL loop.

V. Experimental Setup

To validate the performance of the proposed V2LLM framework, we conduct extensive simulations using a preprocessed real-world V2X telemetry dataset enhanced with synthetic wireless characteristics.

A. Dataset and Telemetry Preparation

We use real vehicle trajectory traces from the HighD [11] and NGSIM [12] datasets, which contain highway and urban driving logs captured at high temporal resolution (10 Hz). For each vehicle v_i, we extract telemetry vectors $\mathbf{x}_i(t) \in \mathbb{R}^{12}$ containing mobility parameters such as (p_x, p_y, v, a, θ) and augment them with synthetic wireless communication features:

- *RSSI*: Computed using a log-distance path loss model based on distance to the nearest RSU.
- *PLR*: Derived using empirical models of wireless interference and density-based congestion.

To simulate communication outages, we inject packet loss events using vehicle-specific Bernoulli masks, where missing rates are sampled from a uniform distribution $\sim \mathcal{U}(0.1, 0.3)$. DT descriptors are computed at each RSU r_j using a sliding window of 5 seconds to estimate $\mu_j(t)$, $\sigma_j(t)$, $\rho_j(t)$, and lane-level occupancy $\psi_j(t)$. Each RSU maintains a retrieval memory $\mathcal{K}_j$ containing $M = 500$ telemetry windows and DT state pairs. The simulation parameters are detailed in Table .

TABLE I: Simulation Parameters

Parameter	Value	Parameter	Value
N	100	R	5
Telemetry Freq.	10 Hz	T	3000
Telemetry Dim d	12	d_z	8
$\lvert\mathcal{H}_i(t)\rvert$	5	λ_1, λ_2	0.01, 0.05
Missing Rate	10–30%	K	4
Sliding Steps n	up to 5	LLM Model	GPT-2 (124M)
LR	10^{-3}	T_{sync}	100 steps

B. Baselines for Comparison

We evaluate the proposed V2LLM framework against the following baselines:

- *Linear Interpolation*: Direct interpolation between neighboring timestamps.
- *KNN Imputation*: Nearest-neighbor-based reconstruction in trajectory space.
- *Seq2Seq*: LSTM-based sequence-to-sequence prediction.
- *FedAvg*: Federated LSTM models trained across RSUs with synchronous averaging.

C. Evaluation Metrics

We report both telemetry recovery and classification performance using the following metrics:

- **Reconstruction MSE (Single-step)**:

$$\text{MSE}_{\text{rec}} = \frac{1}{|\mathcal{M}|} \sum_{t \in \mathcal{M}} \|\hat{\mathbf{x}}_i(t) - \mathbf{x}_i(t)\|_2^2 \tag{18}$$

- **Telemetry Similarity (Cosine)**:

$$\text{Sim} = \frac{1}{|\mathcal{M}|} \sum_{t \in \mathcal{M}} \frac{\langle \hat{\mathbf{x}}_i(t), \mathbf{x}_i(t) \rangle}{\|\hat{\mathbf{x}}_i(t)\|_2 \cdot \|\mathbf{x}_i(t)\|_2} \tag{19}$$

- **Autoregressive Drift MSE (Multi-step)**: To quantify generation drift in multi-step predictions:

$$\text{MSE}_{\text{drift}}(n) = \frac{1}{n} \sum_{\ell=1}^{n} \|\hat{\mathbf{x}}_i(t+\ell) - \mathbf{x}_i(t+\ell)\|_2^2 \tag{20}$$

- **Forgetting Measure (FM)**: To evaluate continual learning stability:

$$\text{FM} = \frac{1}{T} \sum_{t=1}^{T} \left(\max_{\tau < t} \text{Acc}_\tau - \text{Acc}_t \right) \tag{21}$$

D. Simulation Results

Fig. 2 compares telemetry reconstruction performance across baselines. V2LLM achieves the lowest reconstruction MSE and highest cosine similarity, indicating its effectiveness in restoring both the magnitude and directional structure of high-dimensional telemetry vectors. While Seq2Seq and FedAvg-LSTM offer moderate improvements, they struggle under non-i.i.d. and shifting contexts due to limited environmental grounding. Linear interpolation and KNN perform poorly under temporal misalignment and sparse input. By leveraging DT-informed retrieval and structured prompts, V2LLM enables environment-aligned recovery even under multi-step gaps,

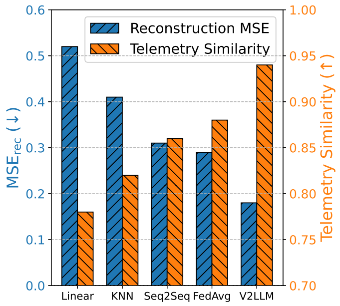

Fig. 2: $\mathrm{MSE}_{\mathrm{rec}}$ and Telemetry similarity.

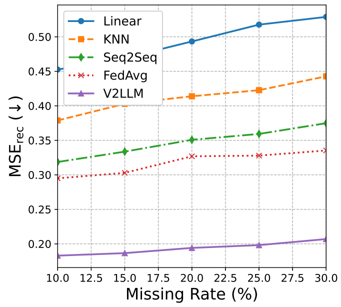

Fig. 3: $\mathrm{MSE}_{\mathrm{rec}}$ vs. missing rate.

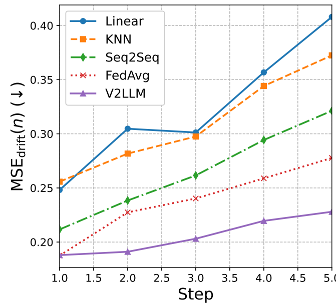

Fig. 4: $\mathrm{MSE}_{\mathrm{drift}}$ vs. prediction step.

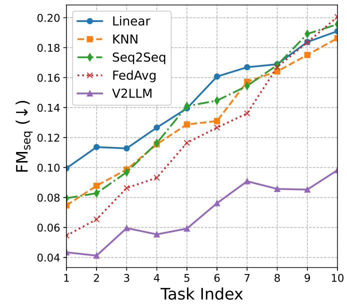

Fig. 5: FM vs. task sequence.

ensuring high-fidelity reconstruction essential for downstream V2X tasks.

Fig. 3 presents reconstruction error $\mathrm{MSE}_{\mathrm{rec}}$ versus telemetry missing rate. As the rate increases from 10% to 30%, all baselines show degradation. Traditional methods exhibit steep error growth due to limited generalization across long gaps. Sequence models like Seq2Seq and FedAvg are more robust but still suffer under data sparsity and distribution drift. In contrast, V2LLM consistently maintains low error, validating the effectiveness of RAG-enhanced prompting and digital twin-guided retrieval in handling severe loss conditions.

Fig. 4 shows $\mathrm{MSE}_{\mathrm{drift}}(n)$ under increasing prediction steps. All methods exhibit growing error due to compounding deviations from autoregressive generation. Linear and KNN degrade rapidly without feedback mechanisms, while Seq2Seq and FedAvg decline beyond 3 steps due to memory limitations and federated drift. V2LLM sustains lower drift across horizons, benefiting from DT-aware prompt construction and RAG-based conditioning, enabling coherent multi-step recovery critical for long-duration telemetry loss.

Fig. 5 tracks the forgetting measure $\mathrm{FM}(t)$ across sequential learning tasks. V2LLM achieves the flattest curve, demonstrating strong retention of past knowledge and minimal degradation over time. This resilience stems from its federated continual learning design, which integrates regularized local updates and global aggregation. In contrast, FedAvg and Seq2Seq exhibit moderate forgetting, while Linear and KNN degrade sharply due to lack of adaptive memory. These results confirm V2LLM's ability to generalize across time-varying telemetry and sustain performance under evolving V2X conditions.

VI. Conclusion

This paper presented *V2LLM*, a unified framework for resilient telemetry recovery and fault classification in V2X networks experiencing intermittent data loss. By integrating digital twin-based environment modeling, retrieval-augmented LLM prompting, and federated continual learning, V2LLM enables accurate reconstruction of missing telemetry and robust, adaptive classification across dynamic traffic scenarios. Through hybrid similarity-based retrieval and structured prompt construction, the framework leverages both spatial-temporal patterns and environmental context to guide autoregressive LLM generation. Additionally, the integration of continual learning at RSUs ensures knowledge retention and adaptability under evolving, non-i.i.d. conditions. Extensive simulations on real-world vehicular datasets show that V2LLM consistently outperforms traditional interpolation, sequence models, and federated baselines across key metrics.

References

[1] Y. Pang, B. Yang, H. Tu, Y. Cao, and Z. Zhang, "Language models can see better: Visual contrastive decoding for llm multimodal reasoning," in *Proc. IEEE ICASSP 2025*, pp. 1–5.

[2] J. Achiam, S. Adler, S. Agarwal, L. Ahmad, I. Akkaya, F. L. Aleman, D. Almeida, J. Altenschmidt, S. Altman, S. Anadkat *et al.*, "Gpt-4 technical report," *arXiv preprint arXiv:2303.08774*, 2023.

[3] M. Hindi, L. Mohammed, O. Maaz, and A. Alwarafy, "Enhancing the precision and interpretability of retrieval-augmented generation (RAG) in legal technology: A survey," *IEEE Access*, vol. 13, pp. 46 171–46 189, 2025.

[4] B. Hazarika, K. Singh, C.-P. Li, A. Schmeink, and K. F. Tsang, "RADiT: Resource allocation in digital twin-driven UAV-aided internet of vehicle networks," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 11, pp. 3369–3385, 2023.

[5] G. Hildebrandt, D. Dittler, P. Habiger, R. Drath, and M. Weyrich, "Data integration for digital twins in industrial automation: A systematic literature review," *IEEE Access*, vol. 12, pp. 139 129–139 153, 2024.

[6] B. Hazarika and K. Singh, "AFL-DMAAC: Integrated resource management and cooperative caching for URLLC-IoV networks," *IEEE Trans Intell Veh*, vol. 9, no. 6, pp. 5101–5117, 2024.

[7] X. Yang, H. Yu, X. Gao, H. Wang, J. Zhang, and T. Li, "Federated continual learning via knowledge fusion: A survey," *IEEE Transactions on Knowledge and Data Engineering*, vol. 36, no. 8, pp. 3832–3850, 2024.

[8] L. Specia (et al.), "Grounded sequence to sequence transduction," *IEEE J. Sel. Top. Signal Process*, vol. 14, no. 3, pp. 577–591, 2020.

[9] S. Gebreab, A. Musamih, K. Salah, R. Jayaraman, and D. Boscovic, "Accelerating digital twin development with generative AI: A framework for 3D modeling and data integration," *IEEE Access*, vol. 12, pp. 185 918–185 936, 2024.

[10] S. Iqbal, X. Zhong, M. A. Khan, Z. Wu, D. A. AlHammadi, W. Liu, and I. A. Choudhry, "Hierarchical continual learning for domain-knowledge retention in healthcare federated learning," *IEEE Trans. Consum. Electron.*, pp. 1–1, 2025.

[11] R. Krajewski, J. Bock, L. Kloeker, and L. Eckstein, "The highd dataset: A drone dataset of naturalistic vehicle trajectories on german highways for validation of highly automated driving systems," in *Proc. IEEE ITSC 2018*, pp. 2118–2125.

[12] M. Khakzar, A. Rakotonirainy, A. Bond, and S. G. Dehkordi, "A dual learning model for vehicle trajectory prediction," *IEEE Access*, vol. 8, pp. 21 897–21 908, 2020.