Privacy-preserving Machine Learning in Internet of Vehicle Applications: Fundamentals, Recent Advances, and Future Direction

Nazmul Islam and Mohammad Zulkernine

Abstract—Machine learning (ML) has revolutionized Internet of Vehicles (IoV) applications by enhancing intelligent transportation, autonomous driving capabilities, and various connected services within a large, heterogeneous network. However, the increased connectivity and massive data exchange for ML applications introduce significant privacy challenges. Privacy-preserving machine learning (PPML) offers potential solutions to address these challenges by preserving privacy at various stages of the ML pipeline. Despite the rapid development of ML-based IoV applications and the growing data privacy concerns, there are limited comprehensive studies on the adoption of PPML within this domain. Therefore, this study provides a comprehensive review of the fundamentals, recent advancements, and the challenges of integrating PPML into IoV applications. To conduct an extensive study, we first review existing surveys of various PPML techniques and their integration into IoV across different scopes. We then discuss the fundamentals of IoV and propose a four-layer IoV architecture. Additionally, we categorize IoV applications into three key domains and analyze the privacy challenges in leveraging ML for these application domains. Next, we provide an overview of various PPML techniques, highlighting their applicability and performance to address the privacy challenges. Building on these fundamentals, we thoroughly review recent advancements in integrating various PPML techniques within IoV applications, discussing their frameworks, key features, and performance evaluation in terms of privacy, utility, and efficiency. Finally, we identify current challenges and propose future research directions to enhance privacy and reliability in IoV applications.

Index Terms—Deep Learning, Internet of Vehicles, Privacy preserving, Intelligent Transportation, Autonomous driving.

I. INTRODUCTION

HE internet of vehicles (IoV) integrates advance communication technologies and intelligent systems to establish a connected ecosystem of vehicles, infrastructure, and services. By enabling real-time vehicle-to-everything (V2X) communication and decentralized computing, IoV enables reliable services for connected

vehicles in smart cities, facilitating efficient traffic management, traffic incident reduction, autonomous driving and enhanced urban mobility [1], [2], [3]. Leveraging real-time data, IoV systems can predict traffic patterns, optimize routes, and dynamically adjust traffic signals for improved transportation efficiency. For autonomous driving, these systems integrate advanced sensing and computer vision, enabling vehicles to perceive their surroundings, make informed decisions, and navigate safely without human intervention. Furthermore, IoV systems integrate with urban ecosystem infrastructure to enable smart services such as intelligent parking, dynamic electric vehicle (EV) charging, and personalized infotainment. They also support a unified transportation network that connects mass transit, personal vehicles, and various mobility solutions.

IoV systems inherently generate vast and diverse amounts of data from vehicles, infrastructure, pedestrians, roadside units, and other sources in the ecosystem. Given the data-intensive nature of machine learning (ML) algorithms and their reliance on large-scale datasets for model training, these data form an ideal foundation for ML applications [4], [5]. This has led to the rapid adoption of ML in IoV applications over the past decade. The integration of ML into IoV has significantly improved existing applications in the ecosystem, while also enabling new capabilities such as autonomous driving and smart EV charging. However, IoV data includes sensitive and private information such as location details, driving patterns, and personally identifiable information, raising significant privacy concerns regarding its collection, use, and potential misuse or breaches [4], [6].

In ML-based IoV applications, various types of attacks can compromise both the training and inference stages of the ML pipeline, as well as the data collection and storage processes [7], [8]. During the training phase, adversaries can launch data poisoning attacks, property inference attacks, or Byzantine attacks, whereas in the inference phase, attacks such as membership inference, model inversion, and model extraction can be executed [8], [9]. All these attacks destabilize the model and undermine the privacy and integrity of both the model and its underlying data. Moreover, attacks targeting communication channels, such as man-in-the-middle attacks, can intercept or alter sensitive information, further compromising data security and integrity. The involvement of third-party services in data collection, model training, or deployment introduces additional vulnerabilities, as

Nazmul Islam and Mohammad Zulkernine are with the School of Computing, Queen's University, Kingston, ON K7L 3N6, Canada.

TABLE I
SUMMARY AND COMPARISON OF EXISTING LITERATURE REVIEWS

	Danas	France of the Ctudy		PPML Techniques		IoV Applications				
Paper		Focus of the Study		T2	Т3	T4	T5	D1	D2	D3
I	[8], [9], [10 <u>]</u> [11], [13]], Survey on privacy in DL, including a review of PPML techniques, taxonomy, methods, and challenges and future direction.	•	-	•	•	•	-	-	-
	[7]	Survey of privacy attacks in machine learning	•	-	-	-	•	-	-	-
anes	[12]	Overview of privacy-preserving distributed optimization and learning	-	-	•	•	•	-	-	-
F) F	[14]	Analysis of differential privacy techniques in cyber physical system	-	-	-	-	•	-	-	•
PM [Sc]	[15]	Privacy-preserving blockchain-IoT integration challenges and privacy issues	-	•	-	-	-	-	-	0
nd T	[16]	Systematic review of differential privacy in deep and FL		-	-	-	•	-	-	-
Surveys on PPML mental and Techn	[17]	Survey of privacy-preserving machine learning with fully HE	-	-	•	-	-	-	-	-
irve	[18]	Survey of privacy-preserving deep learning with SMPC	-	-	-	•	-	-	-	-
Su	[19]	Comprehensive taxonomy and review of privacy-preserving FL	•	-	-	-	-	-	-	-
Surveys on PPML Fundamental and Techniques	[20]	Fundamentals, state of the art, trends, and challenges in decentralized FL	•	-	-	-	-	-	-	0
_	[21]	Surveys privacy-preserving and secure robust federated learning techniques	•	-	•	-	•	-	-	-
јс	[22]	Comprehensive classification of security and privacy vulnerabilities in ITS	-	-	-	-	-	•	-	-
ono	[23]	Offers an overview of ITS security challenges and potential solutions	-	-	-	-	-	•	-	-
d Integration Applications	[6]	Comprehensive survey of machine learning for security in vehicular networks	0	-	-	-	-	•	-	-
nteg opli	[24]	Survey of privacy-preservation techniques in electric vehicles	•	0	0	-	•	-	•	-
I pu	[25]	Security and privacy framework for 6G vehicular networks	-	-	-	-	-	-	-	•
s an IoV	[26]	Survey of secure computation methods based on HE in VANETs	-	-	•	-	-	-	-	•
ssue s in	[27]	Survey of local DP techniques for securing IoVs	-	-	-	-	•	•	-	•
y Is	[28]	Review of privacy-preserving solutions using blockchain for VANETs	-	•	-	-	-	•	0	•
Survey on Privacy Issues and Integration of PPML Techniques in IoV Applications	[29]	Systematic analysis of blockchain-enabled federated learning	•	•	-	-	-	•	0	0
	[30]	Examines blockchain intelligence for IoV, including challenges and solutions	-	•	-	-	-	•	•	•
	[31]	Comprehensive review of FL approaches in vehicles	•	0	-	-	-	-	•	0
irve	[32]	Review of recent applications and open problems in FL for ITS	•	0	-	-	0	•	-	0
Su	[33]	Analysis of FL applications in ITS	•	0	-	-	0	•	-	0
Tl	his Study	Provide an overview of the IoV ecosystem, its privacy concerns, and PPML techniques. Detailed review of recent advancements of PPML in IoV applications and discuss current challenges and future directions.	•	•	•	•	•	•	•	•

Notation: ● covered in the study; ○ covered some aspect of the specific IOV application domain or PPML technique; - not covered in the study. The five PPML techniques are T1: FL, T2: BC-PPML, T3: HE, T4: SMPC and T5: DP. The three application domains are D1: Intelligent transportation and traffic management, D2: Autonomous driving and safety-critical applications and D3: Communication infrastructure and smart services.

mishandling of data or breaches at third-party entities can threaten the entire IoV ecosystem and compromise vehicle security and user privacy on a broader scale [10].

These risks pose significant privacy and security challenges in IoV applications, especially with real-time decision-making. To circumvent these risks, advanced privacy-preserving ML (PPML) techniques such as federated learning (FL), homomorphic encryption (HE), secure multi-party computations (SMPC), and differential privacy (DP) are being thoroughly studied to protect user data and model integrity during various stages of the ML pipeline [8], [9], [10]. Furthermore, blockchain-based PPML (BC-PPML) enhances security and trust in distributed IoV systems while preserving privacy. Therefore, this study provides a comprehensive review of PPML techniques in IoV applications, highlighting privacy challenges and mitigations in this dynamic, data-intensive ecosystem.

A. Related Study

For completeness of the study, we have considered two main areas of related research as summarized in Table. I. First, studies that cover various PPML techniques and second, studies that review privacy challenges in specific IoV application domains or examine specific PPML techniques within those domains.

Several survey studies have presented the theoretical and fundamental aspects of PPML techniques. For instance, [9] provided a comprehensive overview of privacy issues in deep learning (DL), addressing various threats and protection methods, while [10] reviewed privacy-preserving techniques for DL, emphasizing their effectiveness and limitations. Study [8] examined methods, challenges, and future directions in PPML, offering insights into the research trends. Additionally, [7] categorized and analyzed privacy attacks in ML, presenting a detailed survey of various attack methods. In the context of distributed systems, [11] surveyed distributed DL alongside privacy-preservation techniques, and [12] focused on privacy-preserving distributed optimization and learning, discussing advancements and challenges in the field. Furthermore, [13] provided a comprehensive taxonomy and structured overview of privacy-preserving DL. Besides broader survey studies, there are studies that focus on specific PPML techniques. Authors in [14] provided a comprehensive survey on DP techniques for cyber-physical systems, examining their applications across various domains. Similarly, [15] extensively covered privacy preservation in

blockchain-based IoT systems, discussing integration challenges, prospects, and future opportunities. Study [16] focused specifically on application of DP in deep-FL, analyzing its implementation and effectiveness. Addressing secure computation, [17] surveyed PPML using fully-HE (FHE), discussion its potential applications and limitations. Authors in [18] reviewed PPML via SMPC and encryption, summarizing techniques for training and inference. Furthermore, FL has become a key focus and most studied in PPML research. Authors in [19] provided a comprehensive survey on FL, including a taxonomy and future directions, while [20] examined decentralized FL, discussing its fundamentals, current advancements, and challenges. Study [21] addressed privacy and security concerns in collaborative learning by surveying privacy-preserving, secure, and robust FL techniques. While these studies provided key insights into the theoretical foundations and applications of PPML, they do specifically address the unique challenges and requirements in domain-specific applications like in the IoV.

In the domain of IoV, several survey papers provide a broad overview of privacy and security challenges within various applications in the domain. For instance, [22] discussed classification of security and privacy issues in intelligent transportation systems (ITS) and their key challenges. Similarly, [23] provided an overview of secure ITS, highlighting the associated challenges and potential solutions. The authors in [6] presented a comprehensive survey on the role of ML in ensuring security within vehicular networks. In the context of autonomous driving, [24] specifically focused on PPML for electric vehicles (EVs), presenting its challenges and advancements. A broader survey on security and privacy concerns in 6G-enabled IoV is presented in [25]. A few studies focused on specific PPML techniques within IoV applications, such as the ITS and connected autonomous vehicles (CAVs). Authors in [26] surveyed secure computation using HE in vehicular ad hoc networks (VANETs), addressing its applications and challenges. Study [27] focused on local-DP as a mechanism for securing IoV systems, providing a detailed analysis of its applicability. Additionally, [28] examined BC-PPML in IoV, highlighting the integration of blockchain and its implications for privacy. A systematic literature review on blockchain-enabled FL frameworks for IoV was presented in [29]. Study in [30] explores the use of blockchain intelligence for IoV, discussing some aspects of blockchain-based PPML. Focusing on specific applications within IoV, [31] surveyed FL approaches for CAVs, analyzing existing methods and identifying open challenges. Furthermore, [32], [33] reviewed FL applications in ITS, emphasizing recent developments, applications, and future directions.

B. Scope and Contribution

Existing survey literature generally covered either a broad review of PPML technologies, security and privacy in IoV, or the application of specific PPML techniques within narrow domains of the IoV ecosystem, such as CAVs or ITS. The current study provides a comprehensive survey of PPML techniques within IoV applications. First, we provide an overview of the IoV ecosystem and categorize the applications in the ecosystem into three major domains: (1) intelligent transportation and traffic management, (2) autonomous driving and safety-critical applications, and (3) communication infrastructure and smart services. We then analyze the various data types within each domain and their associated privacy challenges. Following this, we review various PPML techniques and compare their performance and applicability. Building on the foundation of IoV applications and PPML techniques, we review the recent advancements in the adoption of key PPML techniques, including FL, BC-PPML, HE, SMPC, and DP, within the three main application domains. Finally, we identify gaps in current research and propose future directions for the integration of PPML in IoV.

The scope of this survey is privacy-preserving techniques employed within the context of ML processes in IoV applications. Notably, our paper excludes studies adopting privacy-enhancing techniques such as data perturbation, anonymization, pseudonymization, k-anonymity, and others, unless they are specifically integrated into PPML for data privacy in IoV applications. Additionally, techniques related to authentication, authorization, and access control, aimed at securing user identities or access to data, are out of scope.

C. Survey Structure

Fig.1 provides a structural organization of the article. The structure of the survey is as follows. Section 2 provides an overview of the IoV ecosystem and the three key application domains withing IoV. It also covers the various data types and their usage within IoV application domains and discusses the associated privacy challenges in IoV applications. Section 3 introduces PPML and reviews standard techniques, such as FL, BC-PPML, HE, SMPC, and DP, while providing the fundamentals of each technique and comparing their approaches. In Section 4, recent advancements in applying these PPML techniques to the IoV application domains are discussed in detail, with focus on their implementation, key features and performance evaluation, along with summary tables. Section 5 identifies recent challenges within PPML techniques and their adoption in IoV applications, offering potential research directions. Section 6 concludes the paper.

II. IOV APPLICATIONS AND PRIVACY CHALLENGES

The IoV is an ecosystem where vehicles, infrastructure, users, cloud platforms and other entities are seamlessly interconnected through advanced technologies. IoV integrates real-time data acquisition, robust communication networks, intelligent data processing, and application-driven services into a cohesive framework. At its foundation, the system harnesses environmental and vehicular data via sensors, actuators and edge devices, which is then transmitted through high-speed, low-latency networks supporting diverse vehicle-to-everything (V2X) interactions [34], [35], [36]. These data streams are dynamically processed through distributed edge-cloud systems and ML-based analytics, enabling real-time

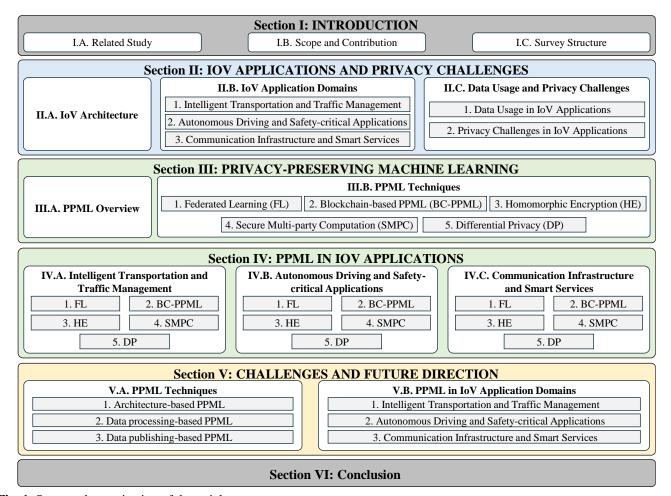


Fig. 1. Structural organization of the article.

decision-making in artificial intelligence (AI) applications such as traffic management, autonomous driving and smart infrastructure and services.

A. IoV Architecture

The IoV architecture consists of multiple layers that work together to enable seamless communication, data processing and applications [3], [37]. Based on the literature review and recent advances in IoV, we have proposed a four-layered IoV architecture illustrated in Fig. 2.

The perception layer in the IoV architecture is responsible for collecting comprehensive data related to vehicles, the environment, and users from various sources, including vehicles, RSU, smart devices, and other connected data points within the network. The data is collected using different sensors and actuators such as global positioning system (GPS), cameras, ultrasonic sensors, light detection and ranging (LiDAR), accelerometers, radars and magnetometers. This data encompasses vehicle-specific metrics such as speed, direction, acceleration, position, and engine condition, as well as traffic-related information like on-road vehicle density, traffic conditions, and weather alerts, alongside multimedia and infotainment records related to users. Therefore, this layer is the principal source of all data within the IoV ecosystem [34], [35], [36]. Furthermore, the perception layer plays a vital

role in the electromagnetic transformation and secure transmission of the data to the subsequent layers, ensuring that the data is digitized and transmitted efficiently and securely.

The communication layer is divided into V2X communication and in-vehicle communication [3], [37]. V2X enables vehicles to exchange information with other vehicles, infrastructure, pedestrians, and the network [36]. It comprises of vehicle-to-vehicle (V2V), vehicle-to-infrastructure (V2I), vehicle-to-network (V2N), and vehicle-to-pedestrian (V2P) communication. The common standards communication are dedicated short range communications (DSRC), wireless access in vehicular environments (WAVE), long-term evolution for vehicles (LTE-V), and fifth generation mobile network (5G)) [1], [36]. DSRC, based on IEEE 802.11p, provides low latency and high reliability for V2V and V2I communication, utilizing onboard units (OBUs), application units (AUs), and roadside units (RSUs)) [1], [36]. LTE-V, introduced in 3GPP Release 14, supports V2X communication through cellular (Uu) and direct (PC5) interfaces, offering higher coverage than DSRC but with higher latency [36]. 5G and 6G technology aims to provide reliable, low-latency transmissions with high bandwidth, supporting integration with other networks, mobile edge computing (MEC), and network slicing to meet the requirements of various AI-based V2X applications [25]. To

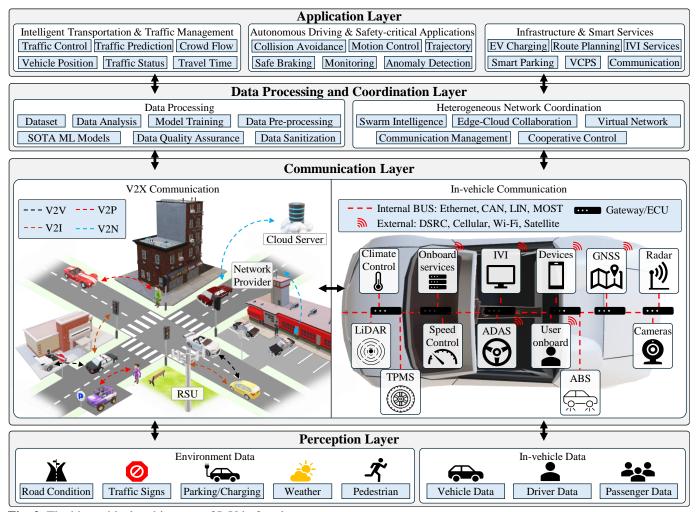


Fig. 2. The hierarchical architecture of IoV in four layers.

address network heterogeneity, this layer processes and standardizes data received from diverse networks, ensuring interoperability. Additionally, it incorporates communication control and management services, enforcing policies like traffic prioritization, load balancing, QoS management, network security, routing optimization, congestion control, and packet inspection to optimize data transmission within the IoV ecosystem.

In-vehicle communication systems integrate both external (wireless) and internal (wired) networks to enable seamless data exchange. External network of V2X leverages communication standards mentioned above to facilitate connectivity with systems such as the global navigation satellite system (GNSS), in-vehicle infotainment (IVI), and advanced driver assistance systems (ADAS). Internal networks enable bus-level communication among electronic control units (ECUs), sensors, controllers, and actuators within the vehicle. These bus-level communication networks can be categorized by their functionality [34], [35]. Controller area network (CAN), local interconnect network (LIN), FlexRay, and intelligent data bus (IDB) are primarily utilized for real-time communication. High-speed data buses such as domestic digital bus (D2B) and media-oriented systems transport

(MOST) facilitate efficient multimedia and high-bandwidth data transfer. Low-voltage differential signaling (LVDS) employs differential signaling for noise-resistant, high-speed communication. The general-purpose peripheral communication interfaces like serial peripheral interface (SPI) and universal serial bus (USB) enable seamless connectivity among microcontrollers, sensors, and external devices. Collectively, these communication networks ensure reliable and high-speed data transfer across diverse vehicular subsystems which are essential for safety-critical functions and high-bandwidth applications such as automatic braking system (ABS) and ADAS [34], [35].

The data processing and coordination layer in IoV architecture is responsible for efficient handling of vast amounts of data generated by connected vehicles, smart devices, and infrastructure. This layer integrates advanced technologies and methodologies, including cloud computing for scalable data storage, big data analytics for extracting meaningful insights, and data management systems for ensuring data integrity and accessibility. In some IoV architectures, the data management layer functions as a centralized information management hub, facilitating the storage, processing, and analysis of data received from lower

layers. However, other architectures employ a more distributed approach, where data management responsibilities are spread across edge and cloud layers to optimize latency, bandwidth, and scalability. Edge servers can handle real-time data processing and decision-making for low-latency tasks, while cloud servers provide scalable storage and powerful computational resources for more intensive analytics and long-term data aggregation. Advanced ML algorithms are extensively employed in the data management layer for real-time processing and analytical tasks, facilitating intelligent applications in the next layer.

The application layer leverages the vast amounts of data collected and processed by the lower layers to enable intelligent applications and services. This layer utilizes advanced ML and DL algorithms to enable a wide range of intelligent IoV applications, including intelligent transportation and traffic management systems, connected autonomous driving, and cyber-physical systems and services. By applying AI algorithms, the application layer can optimize traffic flow, enhance road safety, and provide personalized services to users. For instance, DL models can be trained on historical traffic data to predict congestion and suggest alternative routes in real-time. Reinforcement learning algorithms can be employed to control traffic signals adaptively based on current traffic conditions. Moreover, AIpowered CAVs can communicate with each other and the infrastructure to coordinate their movements, avoid collisions, and improve overall transportation efficiency. Various AIbased IoV applications are discussed in the following section.

B. IoV Application Domains

The IoV is a rapidly evolving ecosystem encompassing diverse applications that can be broadly categorized into three key domains: (1) intelligent transportation and traffic management, (2) autonomous driving and safety-critical applications, and (3) communication infrastructure and smart services. Each domain leverages advanced technologies such as AI, the Internet of Things (IoT), V2X communications, and distributed computing with edge-cloud collaboration. These technologies enable seamless connectivity, real-time data exchange, and intelligent decision-making processes.

1) Intelligent Transportation and Traffic Management: Intelligent transportation system (ITS) encompasses a wide range of applications designed to enhance safety, efficiency, and sustainability in transportation [32]. AI-based computer vision applications, such as object detection, play a pivotal role in ITS by accurately detecting and classifying vehicles, pedestrians, and other road users in real-time [38]. This technology enables advanced collision prevention systems that predict and mitigate potential accidents, enhancing road safety. Traffic management is another critical aspect of ITS, where AI algorithms analyze real-time traffic data to optimize traffic flow, adaptively adjust traffic signals, and provide accurate vehicle positioning utilizing advanced AI algorithms [39]. Additionally, smart public transportation systems utilize AI to optimize routes, predict passenger demand, and improve

overall efficiency, while crowd flow prediction models help allocate resources effectively [40]. Furthermore, AI-driven applications, such as pothole prediction, road damage assessment, and driver misbehavior detection, contribute to enhanced road safety and infrastructure maintenance.

2) Autonomous Driving and Safety-critical Applications: Connected autonomous vehicles (CAVs) integrate advanced communication systems and AI with autonomous driving capabilities to perceive, understand, and navigate complex road environments with minimal human intervention [31]. AI algorithms process and interpret vast amounts of data in realtime for tasks such as object recognition, trajectory prediction, route planning, and cooperative decision-making [41]. CAVs rely on these AI-based perception systems to detect and classify obstacles, pedestrians, and other vehicles by fusing data from multiple sensors and actuators. These systems continuously analyze data to make real-time decisions, facilitate autonomous driving, while ensuring safety and reliability. Safety-critical applications are essential for CAV operations, as they are designed to address scenarios requiring high reliability and precision to avoid collisions and other catastrophic outcomes. Examples include emergency braking systems, collision avoidance mechanisms, and fail-safe operations during sensor or system malfunctions [38], [41], [42]. AI plays a pivotal role in these applications by enabling rapid anomaly detection, redundancy management, and the execution of failover protocols to maintain operational safety in dynamic and unpredictable environments. Reliable communication and coordination between CAVs and other entities on the road through V2X communication are essential for enhancing situational awareness and enabling cooperative decision-making using advanced AI systems [43]. Other applications of AI in CAVs include unmanned aerial vehicle (UAV)-based IoV for traffic management and emergency response [44], advanced driver assistance systems to enhance safety and stable operation [42], and predictive modeling and simulation for vehicle behavior and traffic management [45].

3) Communication Infrastructure and Smart Services: Communication infrastructure and smart services integrate physical and digital systems to enable intelligent, connected, and automated transportation solutions [46]. A key aspect of smart infrastructure is the charging network for electric vehicles (EVs) [47]. AI algorithms optimize charging schedules, predict energy demand, and facilitate smart grid integration, ensuring efficient and sustainable charging operations. Most charging stations (CSs) are strategically located near parking spots to enhance accessibility and convenience. AI-driven parking management systems form another vital component of smart services, leveraging sensors, cameras, and real-time data analytics to guide drivers and CAVs to available parking spots, optimize parking space utilization, and streamline automated payment processes [48]. Smart services also offer personalized user experiences, such as infotainment systems that deliver interactive and tailored experiences to passengers, including real-time information, entertainment, and connectivity services [49]. Personalized

route planning is another AI-enabled service, where algorithms analyze real-time traffic data, user preferences, and contextual information to provide optimized navigation recommendations. These recommendations consider travel time, fuel efficiency, and user constraints, enhancing journey efficiency. Predictive maintenance is another critical application which leverages AI models to analyze sensor data from vehicles and infrastructure to predict potential failures, optimize maintenance schedules, and prevent breakdowns, improving vehicle reliability thereby and reducing maintenance costs. Additionally, usage-based insurance is an emerging smart service, where AI algorithms analyze driving behavior data from connected vehicles to assess risk profiles, offer personalized insurance premiums, and reward safe driving practices, encouraging responsible vehicle usage [50].

C. Data Usage and Privacy Challenges

The IoV relies on diverse data types to enable AI-based downstream applications, enhancing transportation efficiency, safety, and user experience. Few of the common data types in IoV application domains and their privacy challenges are discussed in this section.

1) Data Usage in IoV Applications: The IoV leverages various data types to enhance vehicle functionality, user experience, and the integration of smart services. We have identified eight broad categories of data types that are commonly required for IoV applications and listed as follows. Personal data tracks user travel patterns, frequented locations, and contact information, enabling hands-free connectivity for improved communication and convenience. Biometric data, such as facial recognition and fingerprint scanning, enhances security, personalizes vehicle access, and integrates health monitoring to ensure driver well-being [51]. Behavioral data analyzes driving habits, including speed, braking, and route choices, optimizing safety, enabling applications such as predictive maintenance [50], [52]. Vehicle operation data continuously monitors metrics like engine status and fuel efficiency, supporting reliability and preemptive servicing [53]. Environmental data on weather, road conditions, and cabin climate informs adaptive systems to improve comfort and driving performance [54], [55]. Multimedia and connectivity data store user preferences for navigation, media, and in-vehicle settings while supporting seamless hands-free operations for safety and convenience [49], [56]. Road condition data, collected through vehicle sensors, assesses pavement quality, providing real-time updates to enhance maintenance efficiency and prioritize repairs. Furthermore, IoV requires infrastructure and grid data for EV charging coordination, smart parking solutions, and energy allocation, optimizing resource and enhancing sustainability.

Among the three key domains, intelligent transportation and traffic management mostly leverages personal and environmental data—such as travel patterns, location, weather, and road conditions—to enable advanced navigation, real-time traffic management, hazard warnings, congestion management and dynamic pricing [38], [39], [40]. Autonomous driving and

safety-critical applications depend on all types of personal, behavioral, biometric, environmental and operational data—including vehicle coordination, object recognition, driving habits, vehicle diagnostics, and health monitoring—to support autonomous driving, secure access, adaptive safety systems, advanced navigation and route optimization [38], [41]. Meanwhile, communication infrastructure and smart services mostly utilizes multimedia, infrastructural, environmental, operational and energy data to facilitate smart parking [48], EV charging coordination, predictive maintenance [54], [55], and personalized infotainment [49].

2) Privacy Challenges in IoV Applications: The widespread adoption of AI in IoV applications raises privacy concerns due to the large amounts of data collected, transmitted and stored from vehicles, passengers, infrastructure, and pedestrians [4], [5], [6]. A primary challenge is the exposure of Personally Identifiable Information (PII), including location, driving patterns, and biometric data. Vehicle-specific data, such as make, model, and usage, can also reveal sensitive details, increasing the risk of targeted attacks [22], [23], [24]. Additionally, IoV systems continuously track vehicle movements, exposing information about users' habits and social connections, which is a significant concern for spatial privacy. Data security is another critical challenge, as the transmission and storage of sensitive data create vulnerabilities that may lead to breaches, exposing both individual and aggregate information [11], [12], [20]. Unauthorized access to AI models can also compromise privacy, as algorithms may inadvertently expose personal details through inference, even when data is anonymized. For instance, applications such as route optimization or driving behavior monitoring may unknowingly reveal sensitive information about a driver's routine or frequently visited locations, contributing to behavior profiling and PII exposure.

In the ML pipeline of an IoV application, privacy risks can arise during both the training and inference phases that can compromise the confidentiality and integrity of the system [8], [9], [10]. In the training phase, data poisoning attacks involve adversaries manipulating training data to compromise model integrity or insert backdoors, a significant risk in connected vehicle environments where data quality is crucial for safety. Property inference attacks allow malicious participants to extract statistical properties about others' private training data, potentially exposing vehicle patterns or user behaviors. Byzantine attacks, common in distributed learning scenarios, involve malicious vehicles or data points injecting false gradients during training, destabilizing the model and lowering its accuracy.

During the inference phase, privacy risks are amplified by attacks like membership inference, model inversion, and model extraction [8]. Membership inference attacks enable adversaries to determine whether specific vehicle data was used in training, potentially revealing sensitive patterns. Model inversion attacks aim to reconstruct sensitive training data from model parameters, risking exposure to personal vehicle information, such as location histories. Model

extraction attacks involve attempts to steal the functionality of the model through carefully crafted queries, potentially compromising proprietary vehicle behavior models [8], [24].

The involvement of third-party services for data collection, training, or model deployment further exacerbates privacy concerns. Third parties can mishandle data, expose sensitive information, or become targets of separate breaches, jeopardizing the entire ecosystem. If adversaries access thirdparty data or training pipelines, they could launch targeted attacks (e.g., data poisoning or model extraction) that affect all connected vehicles using the compromised system. Data breaches within third-party services could lead to the loss of proprietary vehicle models or private data, compromising user privacy and vehicle security on a large scale [11], [12], [20]. Additionally, man-in-the-middle or other interception attacks on communication channels between vehicles and third-party infrastructures (such as cloud-edge systems) can intercept or alter data, threatening privacy and security by leaking sensitive information or effecting the model input/output.

III. PRIVACY-PRESERVING MACHINE LEARNING

Privacy-preserving machine learning (PPML) address the privacy concerns associated with ML and DL tasks in IoV applications. As IoV systems collect and process massive volumes of sensitive and private data, the need for PPML solutions has become increasingly crucial to preserve privacy at both user and server side [8], [10]. These solutions aim to protect privacy during the ML processes. PPML techniques integrate various mechanisms into ML pipelines and design privacy-preserving techniques to mitigate risks such as model inversion attacks and data leakage among participants in collaborative learning scenarios [8]. Although implementing PPML solutions presents challenges such as model utility loss privacy, adopting differential and when communication or computation overhead when using SMPC or advanced cryptosystems, PPML remains essential for maintaining user privacy and trust in IoV applications. This section provides an overview of existing PPML techniques, emphasizing the privacy-preserving stages in ML applications.

A comparative analysis of the various PPML techniques is given in Table. II.

A. PPML Overview

PPML operates across three main phases: model generation, model training, and model serving [8]. During model generation, privacy protection is implemented in the initial development and architecture design stages. Model training incorporates privacy-preserving techniques throughout the learning process, while model serving ensures privacy during deployment and inference stages, protecting both user queries and model outputs. The privacy guarantees in PPML can be categorized into two main categories [8]. Object-oriented guarantees focus on protecting specific components, including data-oriented protection for raw training data and user inputs, and model-oriented protection for model parameters and architecture. Pipeline-oriented guarantees are generic end-to-

end privacy protection across the entire ML workflow.

B. PPML Techniques

The focus of this study is on three distinct types of PPML approaches integrated into IoV application domains. Firstly, architecture-based PPML approach representing distributed architectural solutions of federated learning (FL) blockchain-based PPML (BC-PPML). FL enables distributed training while keeping data local to its owners, coordinating model updates across multiple parties, while BC-PPML adds an additional layer of transparency and accountability to the ML process, ensuring secure model updates and verification mechanisms. Secondly, the data processing-based PPML methods include homomorphic encryption (HE) and garbled circuits, where HE enables computation on encrypted data, allowing model training without exposing raw data, and garbled circuits facilitate multi-party computation (SMPC), protecting intermediate calculations during training. Finally, the data publishing approach which consist of only differential privacy (DP) for the review. DP introduces statistical noise to protect individual privacy while maintaining useful aggregate information for model training.

1) Federated Learning: FL is a PPML technique that enables training models on distributed datasets without directly sharing raw data. This method not only enhances data privacy but also addresses challenges associated with data heterogeneity and communication efficiency [8]. As illustrated in Fig. 3, FL can be categorized into two main types, centralized FL (CFL) and decentralized FL (DFL) [20]. In CFL, a central server coordinates the learning process. Participants, also known as clients or nodes, train models locally using their data and send the updates (usually model gradients or parameters) to the central server. The server refines the global model by aggregating these updates using various aggregation algorithms and then sends back the updated global model to the clients for further training, inference, or deployment in applications. Two popular data aggregation algorithms are federated averaging (FedAvg) [57] and federated distillation [20]. FedAvg performs weighted averaging of client updates for data aggregation at the server side using:

$$\theta^{(t+1)} = \sum_{v=V} \frac{\left| \varepsilon_v^{(t)} \right|}{\left| \varepsilon^{(t)} \right|} \theta_v^{(t+1)},\tag{1}$$

where, $\theta_v^{(t+1)}$ are the updated parameters from vehicle v, and $\varepsilon_v^{(t)}$ is the size of its local dataset. Whereas, on the client side, the objective function for the local training of the ith vehicle can be expressed as:

$$f_i(x_i) = E_{\varepsilon_i \sim D_i}[l(x_i, \varepsilon_i)], \tag{2}$$

where x_i represents the model parameters, ε_i is a data sample from the local dataset D_i , and $l(x_i, \varepsilon_i)$ is the loss function. The aggregation forms a global model, which is then distributed back to the participants for further training. This process continues iteratively until the model converges [57].

FedAvg falls short in real-world heterogeneous data

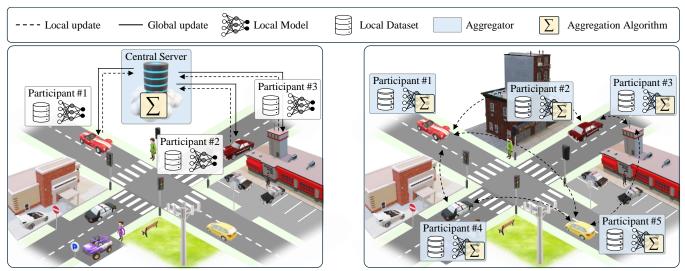


Fig. 3. Basic illustration of centralized FL (left) and decentralized FL (right) structure in IoV.

environments, since a single global model may not suit the individual vehicles, and frequent local updates can result in divergence from the global objective [58]. Knowledge distillation [59], [60] overcome the challenges of model heterogeneity and data heterogeneity by allowing different clients to use varied model architectures (by exchanging logits instead of model parameters), and mitigating non-independent and identically distributed (non-IID) issues in data respectively. In knowledge distillation the knowledge (typically in the form of logits) is transferred from a complex teacher model to simpler student models without sharing raw data. In this process, each client trains its local model in private data and generates logit outputs, which are shared with a central server. The server aggregates these logits to form a global logit, representing the distilled knowledge from the teacher model. Clients then use this global logit to train their local models by minimizing loss function that combines crossentropy loss with Kullback-Leibler divergence, aligning the student's outputs with the teacher's soft targets [60]. The student is trained using linear combination of loss functions:

$$L = (1 - \lambda)L_{CE}(q_s, y) + \lambda L_{KL}(q_s^{\tau}, q_T^{\tau}),$$
 (3) where L_{CE} is cross-entropy loss between the student's predictions q_s and true labels y . L_{KL} is Kullback-Leibler divergence between the student's soft targets q_s^{τ} and the teacher's soft targets q_T^{τ} , λ is a hyperparameter that balances the two loss components. T is a temperature parameter to smooth the probability distribution produced by the SoftMax.

While CFL is effective in various scenarios, it has limitations such as a single point of failure, bottlenecks at the server, and various privacy risks associated with centralized data aggregation [20]. DFL addresses these limitations by eliminating the need for a central server. Instead, each client trains its local model independently and exchanges model updates with its peers (participating nodes). This peer-to-peer communication enhances fault tolerance, robustness against single points of failure and reduces trust dependencies on a central entity. In DFL, achieving consensus among participating nodes is crucial to ensure that model updates are

coherent and reflect the contributions of all peers [20]. A common approach to achieve consensus involves using a mixing matrix $w_{ij}(t)$, where each element w_{ij} represents the weight of communication between node i and node j. The update rule for each peer's model can be expressed as:

$$\theta_i(t+1) = \sum_{i=1}^n w_{ij}(t)\theta_i(t). \tag{4}$$

The number of nodes is given by n and $\theta(t)$ represents the state of either node at time t. The matrix ensures that each peer's update considers information from its neighbors, weighted by the strength of their connection. This balances contributions from different nodes and achieving consensus without a central coordinator. The communication among peers can be organized based on different network topologies, such as fully connected networks, partially connected networks, or clustered networks [20]. In fully connected networks, every node communicates directly with every other node, ensuring high reliability and robustness, but at the cost of increased communication overhead as the network grows. Partially connected networks connect nodes to only a subset of other nodes, often forming structures like star or ring topologies. These configurations reduce communication costs but introduce bottlenecks or increase latency due to the reliance on specific nodes for data transmission. Node clustering involves grouping nodes into clusters based on criteria like geographical proximity or data similarity. Each cluster operates semi-independently but can communicate with other clusters through designated proxy nodes, balancing communication efficiency and scalability while maintaining robustness within each cluster. The choice of topology affects the overall privacy, robustness, flexibility, fault tolerance, and communication costs of the network [20].

2) Blockchain-based PPML: Decentralized and cryptographic nature of blockchain technology address several privacy concerns inherent in traditional, centralized ML architectures. In traditional ML systems, data is often stored and processed on centralized servers, making it vulnerable to

data breaches and unauthorized access. Blockchain provides a decentralized ledger that records transactions securely and immutably. This ledger is maintained across multiple nodes, ensuring there is no single point of failure and no single entity has control over the entire dataset, thus preserving privacy [61], [62]. Blockchain also employs various cryptographic techniques to secure data. Each transaction or data entry on the blockchain is encrypted and linked to previous entries, forming an immutable chain. This ensures that once data is recorded, it cannot be altered without consensus from the network, providing a robust mechanism against tampering and unauthorized modifications [61].

Blockchain consists of several key layers. The data layer stores transaction data using cryptographic methods like elliptic curve cryptography (ECC) for data integrity and confidentiality [15], [62]. In the network layer peer-to-peer communication enables secure data broadcasting and authentication. The consensus layer employs lightweight algorithms such as practical byzantine fault tolerance (PBFT) or directed acyclic graphs (DAG) to achieve agreement on transaction validity without energy-intensive mining processes [61]. Lastly, the application layer supports various IoV applications by utilizing smart contracts to automate processes, prevent unauthorized interventions and enforce rules to ensure secure operations.

Blockchain technology can be combined with other privacypreserving techniques to further enhance privacy in ML systems. Blockchain supports pseudonymity by allowing users to interact with the system through pseudonymous addresses rather than real identities. This feature is crucial for privacypreserving ML, as it prevents the exposure of participants' identities during data transactions and model training processes. Advanced cryptographic techniques such as zeroknowledge proofs (ZKP) and HE integrated into blockchain systems also significantly enhance data privacy [30], [62]. ZKPs allow one party to prove to another that a statement is true without revealing any information beyond the validity of the statement itself. This can be particularly useful in ML for verifying computations or model parameters without exposing underlying data. Furthermore, blockchain can enhance FL by providing a secure and decentralized framework for aggregating model updates from different parties [29]. The blockchain-based federated framework securely aggregates local gradients from untrusted parties using cryptographic techniques, ensuring confidentiality and auditability throughout the collaborative training process. In FL, blockchain can also introduce incentive mechanisms to encourage honest participation and fair behavior among contributors [28]. Leveraging smart contracts and consensus protocols, blockchain ensures that all parties adhere to agreedupon rules, and penalties are enforced for malicious actions.

3) Homomorphic Encryption: PPML utilizes HE to maintain data confidentiality during the training and inference phases of ML pipeline. HE allows computations to be performed on encrypted data without needing to decrypt it, preserving privacy of sensitive user data. HE =

(KeyGen, Enc, Dec, Eval), is composed of four probabilistic polynomial-time algorithms and they are defined as follows. Key generation (HE_{KeyGen}) takes a security parameter λ as input and outputs a public key (p_k) , a secret key (s_k) , and an evaluation key (e_k) , represented as $(p_k, s_k, e_k) \leftarrow HE_{KeyGen}(\lambda)$. Encryption (HE_{Enc}) involves taking the public key p_k and a plaintext message m as inputs to produce a ciphertext c, expressed as $c \leftarrow$ $HE_{Enc}(p_k, m)$. Decryption (HE_{Dec}) requires the secret key s_k and the ciphertext c as inputs to output the decrypted message m^* , denoted as $m^* \leftarrow HE_{Dec}(s_k, c)$. Evaluation (HE_{Eval}) takes the evaluation key e_k , a function f, and a series of ciphertexts c_0, \dots, c_{l-1} as inputs. Here, each ciphertext c_i corresponds to a plaintext m_i for i = 0, ..., l - 1, where l is the number of ciphertexts. The evaluation outputs a final ciphertext c_{fin} , expressed as $c_{fin} \leftarrow HE_{Eval}(e_k, f, c_0, ..., c_{l-1})$ such that $HE_{Dec}(s_k, c_{fin}) = f(m_0, ..., m_{l-1})$. The function f represents an operational circuit over the plaintext space.

HE was first introduced in [63] and since then there has been many developments in various HE algorithms. HE can be broadly categorized into three types: partially homomorphic, somewhat (or strong) homomorphic, and FHE [64]. Partially HE supports unlimited operations of a single type (e.g., addition or multiplication). For instance, the Rivest-Shamir-Adleman (RSA) cryptosystem allows multiplication on encrypted data without decryption, known as multiplication HE. Similarly, the Paillier cryptosystem supports addition and is termed addition HE. The Boneh-Goh-Nissim (BGN) cryptosystem is well-known for supporting both addition and multiplication but only allows a limited number of operations, making it somewhat homomorphic rather than fully homomorphic. The first FHE scheme was proposed using lattice-based cryptography, which introduced the concept of bootstrapping to refresh ciphertexts and manage noise accumulation during computations [64]. This has led to further research into schemes based on lattice theory like learning with errors (LWE) and ring-LWE, as well as those based on approximating the greatest common divisor [64].

Various tools like HE library (Helib), FHE over the ware (FHEW) and HE for arithmetic of approximate numbers (HEEAN) are expanding HE applications in areas such as enhancing cloud computing security. Specifically, in PPML, FHE is a powerful tool because it allows for arbitrary computations on encrypted data, ensuring that both training and inference processes can be conducted securely [65]. For instance, consider a simple linear model represented by the equation $y = w_t x + b$ where w is the weight vector, x is the input feature vector, and b is the bias term. In this context, FHE enables each component to be encrypted: the input as E(x), the weights as E(w), and the bias as E(b). The model can then compute the encrypted output as E(y) = $E(w_t x + b) = E(w_t x) + E(b)$. This ensures that the output remains encrypted until it is decrypted by an authorized party, thereby maintaining data privacy throughout the computation process. Recent advancements have significantly improved the

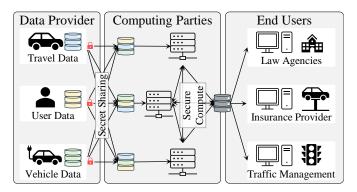


Fig. 4. General illustration of SMPC architecture.

efficiency of PPML using FHE. For instance, HETAL employed HE for transfer learning by encrypting client data using Cheon-Kim-Kim-Song (CKKS) [66]. However, despite these advances, high computational costs and complexity associated with HE still persist in ML and DL integration [64].

4) Secure Multi-party Computation: SMPC can enable collaborative training and inference of ML models without compromising the privacy of individual data inputs as illustrated in Fig. 4. [8], [12], [13]. This is achieved by ensuring that only the final output of the computation is revealed, and individual inputs remain confidential. SMPC protocols are designed to provide input privacy and correctness, ensuring that computations are performed accurately without revealing sensitive information. Input privacy guarantees that no information about private data can be inferred beyond what can be deduced from the output itself. Correctness ensures that even if some parties collude or deviate from the protocol, they cannot force an incorrect result on honest parties. The goal of SMPC is to create a secure protocol that allows multiple participants P_i , where i =1,..., m to jointly compute a function $f(x_1,...,x_m) =$ $(y_1, ..., y_m)$ based on their private inputs x_i [67], [68]. Each participant P_i should receive only their respective output y_i , without gaining extra knowledge, ensuring input privacy.

SMPC leverages several cryptographic techniques such as secret sharing and secure computation to ensure privacy preservation in ML applications [68]. Secret sharing is where data is divided into shares distributed among participants, ensuring that only specific combinations of these shares can reconstruct the original data. This ensures that no single party has access to the complete dataset, maintaining privacy across the board. Whereas, secure computation, like adopting HE, allows computations to be performed on encrypted data, thus keeping data secure throughout its lifecycle. Additionally, techniques such as garbled circuits and oblivious transfer facilitate secure function evaluation by allowing computations on encrypted or obfuscated data without revealing the inputs themselves [68].

In the context of PPML, SMPC enables several critical applications. For instance, it allows for collaborative model training where multiple parties can jointly train ML models using their private datasets without exchanging sensitive information [8], [12], [13]. SMPC also supports secure

inference, enabling parties to perform model inference using private data inputs while ensuring that neither the model owner nor the input provider learns more than necessary about each other's data. Furthermore, SMPC can be applied to feature selection, allowing parties to collaboratively determine important features for a model without exposing raw feature values, enhancing model accuracy while preserving privacy.

5) Differential Privacy: DP provides a mathematical guarantee that individual data points cannot be inferred from the output of a model [14], [16]. It achieves this by introducing controlled randomness into the data analysis process, ensuring that the inclusion or exclusion of any single data point does not significantly affect the outcome. This is particularly important in distributed ML where models are trained across decentralized devices without sharing raw data. In such settings, DP helps protect user data by adding noise to the model updates before they are aggregated at a central server [16]. DP operates on the principle that a privacypreserving mechanism should produce outputs that are statistically indistinguishable whether or not any single individual's data is included [8], [14]. A mechanism M satisfies ε -differential privacy if for any two neighboring datasets d and d' (differing by one individual), and for any possible output subset, S, the probability, Pr, that the mechanism produces an output in S is nearly the same:

$$Pr[M(d) \in S] \le e^{\varepsilon} \Pr[M(d') \in S].$$
 (5)

Here, ϵ is a non-negative parameter that quantifies the privacy loss; smaller values of ϵ indicate stronger privacy guarantees.

To achieve differential privacy, noise is added to the output of a function. This noise is often drawn from a Laplacian distribution, which is determined by the sensitivity of the function being computed. Sensitivity, denoted as ΔQ , measures how much the function's output can change when an individual's data is added or removed:

$$\Delta Q = \max(||Q(d) - Q(d')||_1). \tag{6}$$

Noise drawn from a Laplace distribution is typically added to the function's output to obscure individual contributions:

$$M(d) = Q(d) + Laplace\left(0, \frac{\Delta Q}{\varepsilon}\right),$$
 (7)

here, Q(d) represents the true query result, and Laplace noise ensures that individual contributions are obscured. A relaxed version known as (ε, δ) -DP allows for a small probability δ that the strict privacy guarantee might be violated:

$$Pr[M(d) \in S] \le e^{\varepsilon} \Pr[M(d') \in S] + \delta.$$
 (8)

This relaxation provides more flexibility in designing mechanisms and enhances resistance to attacks using auxiliary information. The Gaussian mechanism, which scales noise according to the L2 norm, is commonly used in this context:

$$M(d) = Q(d) + N\left(0, \frac{(\Delta_2 Q)^2}{\varepsilon}\right),\tag{9}$$

where $N(0, (\Delta_2 Q)^2/\epsilon$ represents Gaussian noise with variance scaled according to sensitivity.

As illustrated in Fig. 5. differential privacy techniques can be broadly categorized into two main approaches: centralized and local DP [27]. In centralized DP, a trusted third party

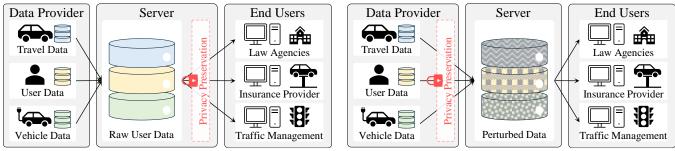


Fig. 5. Illustration of centralized DP (left) and local DP (right) for IoV applications.

TABLE II SUMMARY AND COMPARISON OF VARIOUS PPML TECHNIQUES

		Performance Evaluation				
Technique	Tools/Frameworks	Privacy	Utility	Efficiency		
FL [19]-[21]	TensorFlow Federated; PySyft; FATE; Google FedAvg; Flower; FedML; PFed-HE 32	Data kept locally with minimal raw data exchange; Susceptible to model poisoning and inversion attacks ¹	Near-centralized performance with some accuracy drop ¹ ; Handles non-IID data with bias ¹ ; Convergence and selection biases ¹	High communication overhead ¹ ; Bandwidth and sync constraints ¹ ; Scalability impeded by heterogeneous network/clients ¹		
BC-PPML [15], [29], [30]	Fabric; BE-DPPML; SecureML on	Decentralized trust; immutable ledger; Integration of ZKP; Often combined with HE/MPC/DP for confidentiality; Potential on-chain data leakage if not well designed ¹ ; Public transaction data	Verifiable computations and transparent auditing; Fair contribution tracking; Enhances trust and security in distributed ML settings; Data provenance and auditability	High consensus overhead and storage ¹ ; Energy-intensive ¹ ; Scalability is nontrivial in large networks ¹ ; Computational overhead for consensus and PPML integration ¹		
HE [17], [26]	Microsoft SEAL; Helib; PALISADE; HE-Transformer; TF- Encrypted	Strong cryptographic security guarantees; End-to-end encryption, Data remains encrypted throughout computation; IND-CPA security; Bit- level security; No exposure of plaintext to untrusted servers	Exact results for supported operations; Limited to polynomial operations ¹ ; Supports both training and inference; Fixed-point arithmetic may be required ¹	Extremely high computation overhead ¹ ; Large ciphertext expansion ¹ ; Slow training and inference ¹ ; Complex key management ¹ ; Limited operations ¹ ; limited to polynomial-friendly or carefully designed networks ¹		
SMPC [12], [18]	ABY; MP-SPDZ; Sharemind; SecureNN; FLASH; SecureML; SWIFT 5; CrypTen; SecretFlow – SPU	Information theoretic security; No trusted third party and no raw data revealed to any single party; Strong security under semi-honest; T-out-of-n threshold security	Supports complex operations; Protocol complexity limits adoptability ¹ ; Near-exact results for linear; Complex protocols for non- linear ¹ ; All parties must be active ¹	High communication and computation overhead $^{\downarrow}$; Poor scaling (often $O(n^2)$ or low) $^{\downarrow}$; Efficiency improves with optimized protocols (3/4-party); Protocol round complexity $^{\downarrow}$		
DP [16], [27]	TensorFlow Privacy; OpenDP; PyVacy; PyTorch Opacus; DiffPrivLib	Quantifiable $\epsilon\text{-DP}$ guarantee; Tunable noise addition; Strong theoretical bounds on information leakage	Sharp trade-off with privacy and utility ¹ ; Performance varies with data size ¹ ; Maintains statistical properties with careful noise calibration	Low computational cost; Minimal memory and no extra communication; Overhead scales with dataset size and ϵ calibration $^{\downarrow}$		

Notation: \(\text{Indicate potential limitation of the PPML technique.} \)

collects and processes raw data while adding noise to protect individual privacy during the analysis phase. Whereas local DP allows users to perturb their data locally before sharing it, eliminating the need for a trusted third party and providing stronger privacy guarantees. These techniques can be implemented through various mechanisms in DL, including adding random noise to input samples, gradients, or functions during the training process [14], [16], [27]. Implementing differential privacy involves balancing privacy and utility. More noise generally implies better privacy but degrades model accuracy. Additionally, managing cumulative privacy loss over multiple queries or iterations (the privacy budget) is crucial to maintaining robust privacy guarantees over time.

IV. PPML IN IOV APPLICATIONS

In recent years, there has been extensive research and industrial focus on adopting PPML in IoV applications. As IoV systems handle vast amounts of sensitive data, ensuring data privacy during ML tasks is a major concern, particularly in decentralized infrastructures or collaborative learning

scenarios, as discussed in Section II.C. PPML techniques have been developed to safeguard privacy throughout the ML lifecycle on both the client and server side, as well as during data transit [8], [9], [10]. These techniques counteract both white-box and black-box attacks by embedding privacy-preserving mechanisms directly into the ML pipeline. Furthermore, they mitigate the risk of privacy breaches stemming from third-party services.

From the infrastructure-based PPML, as well as PPML in general, FL has been the most studied and widely researched techniques across various domains, including IoV applications. Meanwhile, most of the literature on BC-PPML focuses on integrating blockchain with other PPML techniques, such as FL, to enable collaborative model training without exposing raw data. It also ensures transparent verification, immutable records, and secure execution of training protocols [29], [30], [69]. Data processing-based (HE, SMPC) and data publishing-based (DP) PPML approaches are also implemented for ensuring the confidentiality of sensitive information, although they are not as widely implemented as

FL in IoV applications. While these techniques often operate independently, hybrid frameworks integrate FL, blockchain, HE, SMPC, and DP to address varied privacy challenges, leveraging their complementary strengths within IoV.

This section reviews recent advancements in integrating PPML techniques into IoV applications, focusing on strategies to enhance privacy at various operational stages. The literature is categorized into the three key IoV domains: (1) intelligent transportation and traffic management, (2) autonomous driving and safety-critical applications, and (3)communication infrastructure and smart services. Within these domains, studies are further classified based on the PPML techniques outlined in Section. III. Some studies adopt a hybrid framework that combines multiple PPML techniques; for clarity and ease of reading, we have included them in all relevant sections corresponding to the techniques employed.

A. Intelligent Transportation and Traffic Management

In the domain of intelligent transportation and traffic Management, real-time data aggregation and analysis are used to optimize traffic flow, reduce congestion, and enhance urban mobility. The integration of various PPML techniques in this domain enables decentralized, privacy-preserving model training and secure data sharing among various IoV entities. In this section, we will review the recent advances in adopting various PPML techniques to address privacy challenges while improving the overall efficiency and reliability of intelligent transportation and traffic management systems.

The technical summary of the literature integrating PPML in intelligent transportation and traffic management systems is given in Table. III.

1) Federated Learning: FL is increasingly being utilized in intelligent transportation systems (ITS) to enhance privacy during downstream traffic management tasks such as mobility flow prediction (traffic or crowd), travel time estimation, traffic signal control and traffic status identification. ML based traffic prediction models, such as graph convolutional Networks (GCNs), have shown robust capability in capturing both spatial and temporal correlations in traffic data due to the inherent graph structure of transportation networks. However, these models often have risks related to data privacy, prolonged training times, and high communication costs, which limit their deployment in real-world [70]. FL addresses these challenges by providing privacy preservation while improving both computation and communication efficiency.

One application of FL for traffic flow prediction proposed by [71] involves replacing the centralized road network model with a decentralized global GCN to balance accuracy and computational cost. Nonetheless, current GCN-based models utilizing FL often neglect the underlying topological structure of the traffic networks, which can lead to potential privacy breaches in ITS. To counter this [72] proposed FASTGNN, a FL approach using GNN for traffic flow prediction that protects graph topological data privacy. The method employs DP by transforming the original adjacency matrix using a Gaussian matrix at the user level and constructing a global

adjacency matrix at the server level. This preserves the privacy of local network structures as the global model is developed. Another development of FL in traffic flow prediction is its ability to handle heterogeneous spatial characteristics across participants using vertical federated learning (VFL) [73]. The authors in [73] introduced a federated graph attention layer to preserve spatial information while capturing short-term temporal features. This allows participants with different spatial attributes to contribute to the global model training without privacy exposure.

Communication efficiency is a critical factor in FL-based traffic prediction systems. In [74], an improved federated averaging algorithm that uses random subsampling of participants has helped reduce communication costs by limiting the number of participants in each round. Additionally, clustering techniques based on geographical proximity (e.g., latitude and longitude) or model similarity have been applied in [75] to further minimize communication overhead. In this approach, clusters of participants share model updates within their group before contributing to the global model, effectively reducing the frequency and volume of communication. An asynchronous communication algorithm proposed in [76] further enhances efficiency by adjusting the aggregation process to account for varying update times among participants.

In addition to traffic prediction, FL is increasingly being utilized in various smart traffic management applications, such as crowd flow prediction, and traffic signal control. The authors in [77] proposed an enhanced FL framework that integrates clustering algorithms to manage human trajectory data. This method extracts spatiotemporal features from the human movement and groups participants with similar characteristics into clusters, thereby improving both the efficiency and accuracy of FL-based crowd mobility predictions. To address the challenge of predicting mobility across various participants with differing locations, a VFL framework for mobility prediction was introduced in [78]. This framework facilitates joint learning over vertically partitioned data from multiple clients, without compromising sensitive location information.

Traffic signal control is another key area where FL has been highly effective. Optimization of traffic signal control [79] has utilized ML techniques like genetic algorithms [80] and particle swarm optimization [81]. However, these approaches do not account for hybrid traffic flow or vehicular control and coordination models. Reinforcement learning (RL) has gained traction in this domain due to its ability to avoid generalized assumptions and complex mathematical operations, but the high dimensionality of the joint action space still poses challenges for centralized RL in large-scale applications [82]. Recent advancements have integrated FL and RL for traffic control systems where agents communicate remotely without routing and loading model parameters during off-peak hours, significantly enhancing convergence speed of the algorithm [83], [84]. Federated deep RL has also been proposed for network edge caching to improve quality of service (QoS) in

wireless networks in [85]. The authors introduced an edge cooperative caching scheme that uses collaborative models (structured as a Markov process) for adaptive caching.

Accurate vehicle positioning in traffic status identification is essential for applications such as navigation, lane-keeping, and collision prevention, which are critical for enhancing road safety and reducing traffic congestion. The authors in [86] employ FL to analyze high-resolution remote sensing images to enhance traffic status identification and congestion monitoring systems while ensuring user privacy. Most of these remote sensing images or other vehicle positioning data are retrieved from widely used global navigation satellite systems (GNSS) and inertial navigation systems (INS). While these systems can provide accurate vehicle positioning, they are costly and heavily reliant on GNSS base stations. Multisystem data fusion from sensor-rich vehicles (SRV) emerged as a potential solution [87], but it remains costly and bandwidth-intensive. Additionally, only a portion of vehicles are equipped with high precision positioning capabilities, limiting widespread adoption. To overcome these challenges, authors in [87] proposed the use of FL to enable cooperative positioning using fused information from both vehicles and infrastructure. This increased training samples by maximizing cooperation with all vehicles and provides high-precision positioning corrections without sharing location data. [87].

Traditional travel mode identification and trajectory prediction used GPS-based segmentation and statistical features for classification [88]. With DL, trajectory data can be mapped into images for convolutional neural networks (CNN) and long short-term Memory (LSTM) to capture spatiotemporal patterns. However, these methods heavily depend on GPS, which is unreliable in low-signal areas. While indoor positioning solutions like Wi-Fi [89] can be utilized, deploying numerous local sensors leads to substantial storage and computational costs. To reduce dependence on GPS data and provide user privacy, authors in [90] proposed an FLbased system where smartphone inertial data is transferred to vehicles to infer the vehicle's position in real-time without GPS signal. This enables real-time position tracking without continuous GPS access, ensuring sensitive data remains localized and user privacy is preserved. A challenge in realworld FL applications for travel mode identification is the lack of labeled data, as vehicle users typically do not label their driving modes, and third-party applications do not collect such labels. To address this, [91] proposed a semi-supervised FL framework that pseudo-labels local data using a small labeled cloud dataset. It aggregates data based on class distribution to handle non-independent and identically distributed (non-IID) issues while keeping user data private on devices.

2) Blockchain-based PPML: Decentralized nature of blockchain enhances privacy in ML frameworks and is often combined with other PPML techniques to improve privacy protection and system robustness. Various studies have integrated blockchain into ML for privacy-preserving ITS applications, including traffic management, traffic congestion control, traffic flow prediction, and destination prediction. For

traffic management, authors in [69] proposed a two-level privacy protection framework incorporating blockchain and DL modules. In the first level, smart contract (SC)-based proof-of-work (PoW) blockchain protocol is utilized to validate data integrity and counteract data poisoning threats. In the second level, a LSTM-autoencoder (LSTM-AE) is employed to transform data into a secure, encoded format, mitigating inference attacks. A blockchain-based FL framework was designed in [92] using a proof of accuracy (PoA) consensus algorithm. This framework resisted Byzantine attacks by ensuring only reliable models are added to the blockchain. This protects privacy and prevents data poisoning or model manipulation.

Authors in [93] proposed a blockchain-based model designed to estimate traffic congestion probabilities while preserving privacy. In [94] a blockchain-enabled FL framework was proposed for urban traffic flow management, enhancing security by verifying model updates through miners to prevent malicious vehicles and reduce data poisoning. It also counteracts inference attacks using local differential privacy in the gradients. In another study [95], BFRT was introduced as a decentralized method for real-time traffic flow prediction, outperforming centralized models. It uses permissioned blockchain technology to protect vehicle privacy while ensuring accurate predictions. Building on to this, a bilevel blockchain architecture was developed for secure FLbased traffic prediction in [96]. It used a distributed homomorphic-encrypted federated averaging (DHFA) approach to secure the federated process and connect decentralized model validation with privacy-preserving measures. Similar to traffic flow prediction, the authors in [97] proposed a blockchain-integrated FL framework for destination prediction, addressing user and location privacy concerns. Besides specific IoV applications, blockchain-based FL can enhance the fairness and privacy of the ML frameworks within IoV systems as a whole. In [98], a decentralized DL architecture using blockchain ensured privacy and fairness through peer-to-peer (P2P) collaboration. It employs a three-layer encryption mechanism combining DP, blockchain, and HE to protect data and improve accuracy, along with a credibility evaluation system to promote fairness and incentivize participants.

In collaborative ML applications, establishing trust among participants and ensuring the credibility of task publishers is crucial. The authors in [99] proposed a hierarchical trust evaluation strategy for 5G-ITS, using a heterogeneous blockchain and ML. It assessed trustworthiness through three levels, with rewards or penalties, and reassessed trust after task completion. This approach encouraged trustworthy behavior and preserved privacy by decentralizing trust management and ensuring transparency and immutability through blockchain. In [100], many-to-one matching model (based on reputation) was proposed to enable secure and efficient assignment of tasks in joint edge learning scenarios. Blockchain is used to securely manage reputation data in a decentralized way, preserving privacy. Smart contracts

automate the training process, ensuring data exposure is task specific. Authors in [101], proposed a two-tier blockchain architecture for FL in mobile edge networks to improve security and efficiency. It used local and global model update chains, with the local chain storing training data in chronological order to build the reputation of local equipment, and the global chain dividing edge nodes into separated and independent chains for specific FL tasks. Privacy preservation was obtained through task fragmentation, decentralized reputation management, and incentives, ensuring minimal data sharing and motivating honest participation.

3) Homomorphic Encryption: HE has been applied to PPML frameworks for IoV applications such as traffic flow prediction, travel time estimation, and traffic signal optimization. In [102], a privacy-preserving traffic flow prediction framework for VANETs used FL, LSTM RNN, and HE to secure model parameters, preventing inference attacks. A prior study [103] introduced a privacy-preserving aggregation scheme for vehicular fog computing using homomorphic threshold cryptosystems. While these studies analyzed security against honest-but-curious fog servers and dishonest users, they did not address advanced inference attacks or evaluate the reliability of secure model updates. In [104], blockchain, ECC, and FHE were used for secure, reliable model updates and privacy-preserving transmission, though ML applications were not considered. Study [96], used blockchain and encryption techniques downstream ML-based traffic prediction task. This framework used HE for privacy preservation and DHFA for secure aggregation of encrypted local updates. It also featured a partial private key distribution protocol and a partial HE/decryption scheme for robust privacy protection.

In ITS, securely processing and aggregating real-time traffic and vehicle data from large number participants enhances predictions and decision-making. To address the privacy and security concerns associated with this data, [105] proposed a fog-based vehicle crowd sensing (FBVC) architecture using HE for secure data collection and aggregation. It features a two-tier fog framework with static upper-tier and dynamic lower-tier fog nodes (fog buses), supporting secure data fusion, traceability, and integrity. Similarly, [106] presented a privacy-preserving sensory data sharing scheme for IoV, leveraging a modified Paillier cryptosystem to ensure location privacy. Using HE, vehicles securely collect and aggregate data, enabling RSUs to perform aggregation without revealing sensitive information. Proxy re-encryption further enables secure querying at the edge offering strong privacy guarantees and resistance to collusion attacks. Additionally, [107] introduced a decentralized location privacy-preserving scheme for spatial crowdsourcing using blockchain and HE. It encrypts vehicle locations with HE to facilitate task allocation, data aggregation, and proximity calculations. It also incorporates order-preserving encryption and non-interactive zero-knowledge proofs (ZKP) to prevent location forgery.

4) Secure Multi-party Computation: IoV applications such as traffic navigation, traffic signal control, and routing

services with location privacy, leverage SMPC to protect sensitive vehicular data while facilitating real-time decisionmaking and collaborative control. The authors in [108] proposed EPNS, a privacy-preserving IoV solution using SMPC and a novel cryptographic construct, multiparty delegated computation (MPDC). By reducing reliance on FHE, EPNS enables efficient secure routing and ML-based traffic prediction while protecting user data. It employs two non-colluding servers—one for re-encryption and another for computation—to prevent data leakage even if one is compromised. A privacy-preserving adaptive traffic signal control framework for connected autonomous vehicles (CAV) was proposed in [109]. The framework integrates SMPC and DP to protect CAV data against three types of privacy threats: collusion attacks, control center database attacks, and inference attacks. The framework uses a linear programming model and arrival rate estimator based on aggregated data, ensuring efficiency in varying traffic conditions. A two-stage stochastic programming model mitigates DP-induced noise impacts on control performance.

While SMPC-based cryptographic protocols provide privacy-preserving computation, leveraging distributed edge architectures can enhance scalability and computational efficiency. The SPEED framework in [110] distributed data processing across edge nodes, reducing centralized attack risks. Using compressed sensing, data shuffling, and split computation, SPEED minimized raw data exposure and enhanced privacy in large-scale IoV. Expanding the application of dual-layered privacy frameworks like in [109] and [110], authors in [111] propose a privacy-preserving decentralized routing service to secure vehicular trajectories using SMPC and DP. It combines additive and Shamir secret sharing to protect location data while enabling accurate traffic estimations. The DP Laplace mechanism enhances privacy in sparse data scenarios, mitigating risks of tracing individual trajectories while ensuring minimal impact on routing.

5) Differential Privacy: Key IoV applications with DP adoption in the literature include traffic flow prediction, realtime traffic monitoring, and location privacy. In ML-based traffic management, local DP (LDP) can be applied to gradient updates in large IoV networks to safeguard vehicle data and prevent attackers from inverting shared gradients. Authors in [112] proposed a privacy-preserving approach for traffic flow prediction by combining FL and LDP. Their method applies Gaussian noise to the gradients of an LSTM model before sharing them with a central server for aggregation, preventing inference attacks. Furthermore, predicting and managing traffic flow effectively requires robust monitoring systems while ensuring privacy protection throughout the process. Authors in [113] presented a distributed traffic monitoring system that protects individual privacy using DP. They introduce three algorithms with different trade-offs between noise levels and computational complexity, with the third approach balancing runtime and noise, making it suitable for real-time traffic monitoring in urban systems.

TABLE III

$SUMMARY\ OF\ EXISTING\ LITERATURE\ ADOPTING\ PPML\ IN\ INTELLIGENT\ TRANSPORTATION\ AND\ TRAFFIC\ MANAGEMENT\ SYSTEM$

Ref.	Application and Data Type	Framework and Key Features	Performance evaluation			
1. Fea	Federated Learning:					
[71]	Application : Traffic flow prediction; Data Type : Traffic flow sensor data; Dataset : PeMS04 and PeMS08 dataset	Framework: FL-GCN integration; Features: Community detection, Local GCN training, Parameter aggregation	Privacy: Data locality, Model privacy; Utility: High accuracy; Efficiency: Reduced overhead with low time cost			
[72]*	Application: Traffic speed forecasting; Data Type: Traffic speed data (sensors); Dataset: PeMSD7 dataset	Framework : FASTGNN; Features : Differential privacy matrix, Adjacency protection, Topology preservation	Privacy: Network and model security; Utility: Satisfactory forecast accuracy; Efficiency: Robust against time-series fluctuation			
[73]*	Application: Traffic flow prediction; Data Type: Urban traffic flow data (sensors, loops); Dataset: TaxiNYC and TaxiBJ dataset	Framework : FedSTN; Features : RLCN module, AMFN module, HE	Privacy : Spatial and temporal privacy; Utility : Improved prediction accuracy; Efficiency : Edge optimization			
[74]	Application : Traffic flow prediction; Data Type : Multi-source traffic data; Dataset : PeMS dataset	Framework: FedGRU; Features: GRU neural network, Federated averaging, Secure aggregation	Privacy : Local data, Parameter privacy; Utility : Minor accuracy dip (0.76 km/h error); Efficiency : Reduced overhead, improved scalability			
[75]	Application: Graph-based traffic forecasting; Data Type: Traffic data (adjacency info); Dataset: PeMSD4 and PeMSD7 dataset	Framework: Graph-based FL; Features: Clustering optimization, Two-step strategy, PSO algorithm	Privacy : Data protection, Model privacy; Utility : Satisfactory accuracy; Efficiency : Reduced communication overhead			
[76]	Application: Traffic state estimation; Data Type: IoV traffic data (vehicular signals); Dataset: England Freeway dataset	Framework: FedTSE; Features: LSTM-based prediction, DRL optimization, Edge computing	Privacy: Data security, State privacy; Utility: State accuracy; Efficiency: Resource optimization			
[77]*	Application: Crowd flow prediction during epidemics; Data Type : Crowd/traffic flow data; Dataset : Generated dataset	Framework : Privacy-aware CFPF Framework; Features : Multi-Factors CNN-LSTM, Clustering algorithm, LDP	Privacy : DP guarantee, Gradient privacy; Utility : Improved accuracy; Efficiency : Reduced communication			
[78]	Application: Mobility forecasting; Data Type: Heterogeneous mobility datasets (split features); Dataset: New York City and Yelp dataset	Framework : VFL framework; Features : Vertical partitioning, Joint domain learning, Neural network algorithms	Privacy : Data partition privacy, Cross-org security; Utility : 4-12% improvement compared to baseline; Efficiency : Resource optimization			
[83]	Application: Adaptive signal control; Data Type: Intersection signal data; Dataset: Generated dataset	Framework: Distributed MARL; Features: Federated averaging, A2C algorithm, Adaptive control	Privacy: Model privacy, Control security; Utility: Enhanced control; Efficiency: Better optimality and learning efficiency			
[84]	Application: Multi-intersection traffic signal control; Data Type : Intersection traffic data; Dataset : Simulated in Cityflow dataset	Framework: FedLight; Features: Multi- intersection control, Federated RL, Autonomous control	Privacy : Local training, Model integrity; Utility : Superior performance; Efficiency : Resource optimization			
[85]	Application: Edge caching for traffic control; Data Type: Usage patterns, Content requests; Dataset: MNIST and Movielens 1m dataset	Framework: Edge caching FL; Features: MEC integration, Cache optimization, Intelligent connectivity	Privacy : Edge privacy, Vehicle security; Utility : High hit rate; Efficiency : Reduced latency			
[86]	Application: Traffic congestion monitoring; Data Type: Remote sensing images; Dataset: Los Angeles and Washington Road dataset	Framework : FL-based monitoring; Features : Distributed learning, Vehicle identification, Realtime monitoring	Privacy : Local data privacy; Utility : 85% detection rate; Efficiency : Low latency (0.047s processing time)			
[87]	Application: Cooperative vehicle positioning; Data Type: Trajectory/location data, sensor-rich vehicles; Dataset: Didi Chengdu dataset	Framework: FedVCP framework; Features: V2V/V2I communication, Transfer learning and data augmentation, Edge computing integration	Privacy: Position privacy, Trajectory protection; Utility: High accuracy and convergence speed; Efficiency: Lower computation and overhead			
[90]	Application: Vehicle Tracking; Data Type: accelerometer/gyroscope data; Dataset: Didi Beijing and Shanghai dataset	Framework: VeTorch; Features: GPS-free tracking, Smartphone sensors, Position inference	Privacy: Location privacy, Sensor security; Utility: lowest MAE; Efficiency: Real-time tracking, high storage (2.25MB)			
[91]	Application: Travel mode identification; Data Type: GPS trajectories; Dataset: GeoLife GPS dataset	Framework : Semi-supervised FL; Features : GPS trajectories, Label propagation, Federated architecture	Privacy : Trajectory privacy, Mode confidentiality; Utility : 90% accuracy with 50% labeled data; Efficiency : Lower communication			
2. Blo	ckchain-based PPML:					
[69]	Application : ITS intrusion detection; Data Type : Intrusion detection and ITS data; Dataset : ToN-IoT and CICIDS-2017 dataset	Framework : Blockchain-enabled DL; Features : Smart contract-based, LSTM-AE encoding, A-RNN for intrusion detection, Enhanced-PoW	Privacy: Data integrity, Inference attack prevention; Utility: Over 98% accuracy; Efficiency: Low overhead			
[92]*	Application : Edge computing; Data Type : Edge device model updates; Dataset : MNIST dataset	Framework : Byzantine-resistant FL; Features : Parallel verification, PoA consensus, Byzantine detection	Privacy: Attack resistance, model verification; Utility: Model integrity, Comparable accuracy; Efficiency: Fast convergence, Lightweight model			
[93]	Application : Traffic estimation; Data Type : Traffic flow data; Dataset : Generated dataset	Framework: Blockchain-DNN; Features: Revenue model incentives, Smart contract prediction, Secure crowdsourcing, PoA consensus	Privacy: User privacy; Utility: High accuracy; Efficiency: High user participation			
[94]*	Application: Traffic prediction; Data Type: Traffic flow data; Dataset: PeMS dataset	Framework: Blockchain-FL; Features: Distributed model updates, Privacy-aware aggregation, Local differential privacy, dBFT	Privacy: vehicle data privacy, defend poisoning attack; Utility: Robustness against attacks; Efficiency: High computation			

(Continued)

[95]	Application: Real-time prediction; Data Type: Traffic flow data; Dataset: DelDOT dataset	Framework : BFRT framework; Features : Edge computing integration, Hyperledger fabric implementation, RAFT consensus	Privacy: Real-time privacy; Utility: Superior accuracy; Efficiency: High throughput and low latency
[96]*	Application : Traffic prediction; Data Type : Traffic prediction data; Dataset : DelDOT dataset	Partial private key distribution, Two-layer	Privacy: Parameter privacy; Utility: Improved performance for controlled group size; Efficiency: Computationally expensive
[98]*	Application: Fair FL; Data Type: Distributed image data; Dataset: MNIST and SVHN dataset	Framework : FPPDL framework; Features : Three-layer encryption, Credibility evaluation, Contribution-based models	Privacy : Update privacy, Model fairness; Utility : High accuracy, Low utility cost; Efficiency : Fair distribution, low communication cost
[99]	Application : Hierarchical trust; Data Type : 5G-enabled ITS data; Dataset : Foursquare dataset	Framework: BHTE strategy; Features: Federated deep learning, Incentive mechanisms, Trust verification	Privacy : Trust verification; Utility : Reasonable and fair trust evaluation; Efficiency : High throughput, Low latency
[100]	Application: Edge learning; Data Type: Edge learning task data; Dataset: MNIST dataset		Privacy : Worker reliability, training integrity; Utility : Optimal matching; Efficiency : Resource optimization
[101]*	Application: Mobile edge FL; Data Type: Edge network data; Dataset: MNIST dataset		Privacy: Model privacy, Long-term reputation; Utility: High accuracy; Efficiency: Reduced delay
3. Hon	nomorphic Encryption:		
[102]*	Application: Traffic flow prediction; Data Type: Traffic flow data in VANETs; Dataset: PeMS dataset	Framework: FL framework; Features: LSTM-RNN model, Secure parameter aggregation, Distributed training	Privacy : Data locality, model privacy; Utility : Train loss below 0.003; Efficiency : N/A
[103]*	Application : Vehicular Fog Navigation; Data Type : Vehicular navigation data; Dataset : Generated	Framework: Privacy-aware FL; Features: Homomorphic cryptosystem, Dynamic resource allocation, Bounded Laplace, Skip list structure	Privacy: Malicious detection, Model parameter privacy; Utility: Flexible participation; Efficiency: Improved computation
[104]*	Application: Decentralized VANET; Data Type: VANET traffic and communication data; Dataset: Generated	Framework: FHE with blockchain; Features: FHE, End-to-end encryption, Smart contract verification, RAFT consensus	Privacy: Model credibility; Utility: Acceptable accuracy, Attack prevention; Efficiency: Improved throughput and lower overhead
[105]	Application : IoV crowd sensing; Data Type : Crowd sensing data; Dataset : Generated from SUMO	Framework: Fog-based security; Features: Multi-level security, Two-tier fog computing, Distributed processing	Privacy: Data protection, Access control; Utility: High throughput in dense traffic; Efficiency: Reduced time, overhead and storage
[106]	Application : IoV Framework; Data Type : IoV sensory data; Dataset : Generated dataset	Framework: Sensory data sharing; Features: Secure sharing protocol, Privacy preservation, Access control, Modified Paillier cryptosystem	Privacy: Data confidentiality, Location privacy; Utility: Low data querying failure; Efficiency: Low computation and communication
[107]*	Application : IoV Crowdsourcing; Data Type : IoV spatial crowdsourcing data; Dataset : Gowalla dataset	Framework: Decentralized Privacy; Features: Circle-based verification, Grid-based location, Multi-level privacy	Privacy: Location privacy, Task confidentiality; Utility: Multi-level privacy protection; Efficiency: High computation
4. Secu	re Multi-party Computation:		
[108]	Application: Cloud VANETs, Optimal routing; Data Type: VANET navigation data; Dataset: Simulation data	Framework: EPNS framework; Features: MPDC encryption, Privacy-preserving navigation, Cloud computation	Privacy: Location, velocity, Navigation privacy; Utility: Optimal path selection; Efficiency: Constant-time encryption
[109]*	Application: Signal Control; Data Type: Connected vehicle traffic data; Dataset: Generated from SUMO	Framework: Adaptive control; Features: Privacy-aware control, Real-time adaptation, Signal optimization	Privacy: Vehicle privacy, Control security; Utility: Low impact on control performance; Efficiency: Low residual queues and delay
[110]	Application: Distributed data processing in ITS Data Type: ITS sensory and traffic data; Dataset: Generated data	Framework: SPEED framework; Features: Multi-level edge computing, LS-SVM modeling, CNN-based detection	Privacy : Data privacy, Compression security; Utility : 99%+ precision, recall F1; Efficiency : Reduced overhead, 94.96% packet delivery ratio
[111]*	Application: Secure vehicular trajectory; Data Type: Vehicle location data; Dataset: TNTP dataset	Framework: Private location; Features: Decentralized routing, Location protection, Laplace mechanism	Privacy: Location obfuscation, Route anonymity; Utility: Accurate travel time estimation; Efficiency: Reduced overhead
5. Diffe	erential Privacy:		
[112]*	Application: Traffic Flow; Data Type: Highway traffic flow data; Dataset: PeMS dataset	Framework : FL-LDP framework; Features : Local differential privacy, Hybrid architecture, Deep neural networks	Privacy: Data protection; Utility: Accurate prediction, Lowest ASR of 2%; Efficiency: Balanced overhead with privacy budget
[113]*	Application : Traffic Monitoring; Data Type : Smart city traffic data; Dataset : Generated dataset	Framework : FL with LDP; Features : Secure aggregation, Local differential privacy, Distributed learning	Privacy: Communication privacy; Utility: Distributed Traffic monitoring utility; Efficiency: High runtime overhead
[114]	Application : EV location privacy; Data Type : EV charging and location data; Dataset : Beijing public charging posts dataset	Framework: Quadtree-based DP; Features: Spatial decomposition, Random sampling, Sparse vector technique	Privacy: Location protection, Charging privacy; Utility: Data availability, 99% query accuracy; Efficiency: Balanced noise
[115]	Application: Data streaming in connected vehicles; Data Type : IoV data; Dataset : TAPASCologne dataset	Framework : IoV privacy; Features : Edge computing integration, Correlated noise addition, Temporal privacy protection	Privacy: Data and connection privacy; Utility: high utility on correlated data; Efficiency: Resource optimization, High computation
		the state (Helevill DDMI)	

Notation: * Indicate integration of more than one PPML in the study (Hybrid PPML).

Besides traffic management, location privacy is crucial for protecting personal identification and routes. Authors in [94] implemented a local differential privacy to protect location data in traffic flow prediction. For EV location data in vehicleto-grid (V2G) networks, a privacy-preserving mechanism using DP, quadtree spatial decomposition, and Bernoulli random sampling was proposed in [114]. Their approach effectively protects location privacy while minimizing relative errors in data utility. Study in [115] developed a DP data streaming system for connected vehicle networks. The study addressed the challenges of data correlation and dynamic network topology by implementing group-based data compression and adaptive noise addition. This method balances location privacy protection with computational efficiency in untrusted vehicular networks where vehicles can join or leave different traffic zones.

B. Autonomous Driving and Safety-critical Applications

Autonomous driving and safety-critical applications require ultra-reliable and low-latency decision-making capabilities. These capabilities are powered by ML algorithms trained on vast datasets from in-vehicle communication and vehicle-to-everything (V2X) networks, as shown in Fig. 2. PPML techniques are essential for ensuring privacy and mitigating adversarial attacks in autonomous driving systems, while also enabling efficient and reliable operation. In this section, we examine recent advancements in the adoption of various PPML techniques to improve privacy and safety in autonomous driving and critical safety applications.

The summary of studies adopting PPML in autonomous driving and safety-critical applications is given in Table. IV.

1) Federated Learning: Connected autonomous vehicles rely heavily on real-time data processing and decision-making to ensure safety and efficiency. FL plays a pivotal role in supporting advanced monitoring systems, including in-vehicle, driver, passenger, and steering-wheel monitoring. Additionally, FL enables privacy-preserving collaboration across vehicles and infrastructure for tasks such as collision avoidance, safe braking, trajectory prediction, and anomaly detection. Furthermore, it facilitates advancements in computer vision tasks, such as object detection, license plate recognition, traffic sign detection, and damage assessments, which are critical for maintaining safety and performance.

In-vehicle monitoring systems enhance service quality and safety across transportation modes by tracking passengers, detecting falls, and identifying emergencies in real time [116]. Given the sensitivity of personal data, these systems must prioritize privacy preservation. FL can ensure individual privacy while enabling decentralized knowledge sharing, enabling models to learn generalizable patterns despite the low frequency of certain events, such as accidents [117]. Various studies have proposed in-vehicle driver monitoring systems to detect driver distractions and provide safety-critical alerts when attention is diverted [45], [118], [119], [120]. However, these systems face significant computational and communication challenges. The FL approach proposed in

[121] utilizes the FedGKT framework [122] to improve bandwidth efficiency through asynchronous training for driver activity recognition, achieving competitive results in both centralized and decentralized model architectures. Furthermore, driver-related data is often tied to personal habits, cultural nuances, and emotional states, making generalization across individuals inherently challenging. To address this, personalized FL technique in [123] enhance model adaptability to diverse driver behaviors, ensuring secure data handling while considering individual driver patterns and requirements. The authors in [124] integrated HE scheme with FL to add an extra layer of privacy in detecting driver drowsiness. Beside driver monitoring, passenger monitoring in public transit is crucial for detecting boarding intent, exit behavior, and risky actions to enhance operational efficiency and safety [125]. However, there is limited research due to data scarcity and challenges in monitoring crowded environments. Nevertheless, advances like DFL framework and gossip protocols in [126], can improve model training in dynamic multi-user environments like public transit. Another in-vehicle monitoring system is steering wheel prediction, which is vital for self-driving and advanced driver assistance systems (ADAS) features like lane-keeping and departure alerts. It estimates steering angles from road images to maintain vehicle alignment, especially on challenging terrains [127], [128]. The authors in [129] used multi-modal data, including road images and optical flow, to enhance accuracy of steering angle prediction models. Another study [130] achieved accuracy similar to centralized models while optimizing edge model quality by adding noise and incorporating various data sources across vehicles. The study also confirms that FL minimized communication overhead and achieved robustness to network disruptions.

The motion controller of a CAV allows a specific trajectory by managing various control aspects, including the accelerator pedal and brake (for longitudinal acceleration/deceleration) and the steering mechanism (for lateral movement) [131], [132], [133]. FL is increasingly used in CAVs for collaborative training and improved controller parameters while ensuring data privacy. The proposed scheme in [134] allowed CAVs to update controller parameters in real time, and continuously improved target speed achievement and vehicle handling through aggregated data. It also allowed flexible vehicle participation, enabling CAV control optimization regardless of the number of participants during training. In collision avoidance and safe braking applications, FL has proven effective in optimizing collaborative control parameters among multiple CAVs at intersections, enhancing collision avoidance significantly without compromising privacy [135]. Similarly, study in [136] used FL for safer braking actions by improving the estimation of road friction coefficients in varied driving conditions and environments. RL for motion control in CAVs is a wellresearched area, mostly because RL can handle complex, dynamic environments by learning control policies through user feedback and sensor data [137], [138]. However, these

systems lacked privacy-preserving features.

Accurate vehicle trajectory prediction enables CAVs to plan movements, anticipate risky behaviors, and prevent accidents. Training these trajectory prediction models rely heavily on spatiotemporal data, including variables like location, speed, and acceleration. Recurrent neural networks (RNNs) and transformers are commonly used for capturing these time series patterns and have been effective in predicting complex movement patterns, both for vehicles and pedestrians. FL framework are capable of learning nuanced spatiotemporal features in combination with transformers [139] and long short-term memory (LSTM) models [140], enhancing predictive accuracy while preserving privacy of vehicle data. Additionally, FL has also been used for anomaly detection in vehicular trajectories. In [141], FL was integrated with the one-class support vector machine (OC-SVM) algorithm anomalous driving behavior detection at intersection points and identify unsafe maneuvers. This approach supported continuous adaptation while minimizing dependency on labeled data, improving detection accuracy while preserving privacy. Furthermore, various studies [142], [143], [144] showed that FL performs comparably to centralized learning while preserving privacy throughout the learning process. To further optimize trajectory prediction, authors in [145] proposed personalized FL which moderated the update frequency to prevent overfitting, enhancing generalization in diverse driving environments. Meanwhile, [146] employed a map fusion in three stage (density-based spatial clustering, score-based averaging, and intersection-over-union-based pruning) to enable accurate, privacy-preserving trajectory predictions without centralizing user data.

Object detection in IoV systems is crucial for safe autonomous driving, relying on advanced computer vision for accurate vehicle and obstacle detection [147], [148]. Algorithms, such as YOLO [149], provides fast, one-stage detection, while R-CNN [150] offers higher accuracy through a two-stage process at the cost of high computation. Studies continue to research optimization strategies that balance the trade-off between detection accuracy and inference speed [151]. FL enhances computer vision tasks in IoV systems by enabling decentralized data processing, preserving privacy, and reducing communication overhead while improving detection through collaborative learning across vehicles [152]. The study in [153] demonstrated that FL architectures enhance object recognition at image boundaries by sharing knowledge (aggregated model weights) across vehicles. For instance, collaboratively trained YOLO improved detection of distant objects like trucks and pedestrians using indirect data from other users. Authors in [154] proposed YOLO-CNN for CAVs to improve safety in snowy conditions. Using FL, the system enabled collaborative training without sharing raw data, preserving privacy while enhancing object detection in challenging winter conditions. Additionally, research in [155] introduced a decentralized FL framework for object classification in CAVs using LiDAR data. PointNet model parameters are exchanged over V2V networks, reducing data

centralization and preserving privacy.

Besides objects, pedestrians and climate, poor road conditions can also pose serious risks to autonomous driving, traffic safety, and vehicle integrity. Authors in [156] used FL for pothole detection, combining 3D-FL and YOLO for accurate defect size estimation. This model could also accurately distinguish real potholes from patched areas and artificial road bumps, countering adversarial ML attacks while preserving privacy. Similarly, [157] proposed a privacy-preserving adaptive FL framework for detecting hazardous road damage, classifying it by severity. Additionally, [158], [159], explore various road surface condition and damage detection methods using distributed FL.

Object detection applications, such as license plate detection/recognition (LPR) and traffic sign recognition, are also essential in CAV systems as these have higher quality dataset for the downstream ML tasks. LPR involves detecting the plate and recognizing its characters, often using region-ofinterest, color, or pixel-to-pixel based methods. However, challenges such as multi-directional detection and motion blur remain in these applications. The authors in [160] redesigned the LPR model for edge devices using privacy-preserving FL, addressing issues such as blur and orientation discrepancies caused by vehicle and sensor dynamics and introduces a tilt correction algorithm of license plates to enhance the model resilience. For traffic sign recognition, the study in [161] proposed FL with a spike neural network based on receptive fields that significantly reduced computational overhead while providing better immunity against noise and improved accuracy compared to traditional CNN-based FL approaches. It also mitigates the risk of location-based privacy leaks inherent in these systems.

To reduce computation overhead and improve FL performance for object detection and other applications, studies have proposed various techniques such as multistage resource allocation and strategic vehicle selection to optimize resource use and network bandwidth [162]. Some techniques address the challenge of limited labeled data, like the semisupervised FL method for image recognition in [163], that used active learning and dynamic hyperparameter tuning to enhance model accuracy in internet of drones. Selective model aggregation proposed in [164], also improved performance by integrating high-quality models from client's with sufficient resources. To further optimize model selection authors in [165] introduce an asynchronous federated aggregation protocol specifically for target recognition. Similar to [164] this protocol also selects optimal local models based on data quality and client's computational capacity, enhancing efficiency and effectiveness of FL model.

2) Blockchain-based PPML: BC-PPML in CAV applications ensures secure and decentralized data sharing and computation. Authors in [166] introduced a collective learning architecture where each CAV can independently train local ML models for tasks like route optimization and emergency response. The distributed nature of this architecture allows CAVs to download validated models from the blockchain,

facilitating efficient collaboration without compromising privacy or system reliability. The authors further combine ML with expert systems and deep RL in study [167] to enhance decision-making capabilities in autonomous driving scenarios, accelerating model convergence and enabling CAVs to adapt to unpredictable situations by leveraging on-board blockchain technology. Study [168] proposed a blockchain-integrated ML framework for vehicle positioning in CAVs. It uses a deep neural network to predict GPS errors and stores the corrections on a blockchain for secure, collaborative error correction, enhancing accuracy while protecting location privacy.

BC-PPML enhances data protection and enables secure collaborative learning while reducing the risk of exposure to malicious entities. Study [169] proposes a blockchainsupported FL framework to detect and mitigate malicious behaviors by distributing ML models across devices and securing training with blockchain consensus. Beside malicious vehicle behaviours, mitigating blockchain integration also ensured transparency and trustworthiness. Study [170] leverages blockchain to incentivize honest participation and balance vehicular privacy with accountability using traceable identity-based schemes. The dual encryption mechanism ensured anonymous authentication for semi-honest vehicles and used blockchain-based traceability and reputation incentives to deter malicious behavior. Similarly, [171] introduces a blockchain-based FL framework to detect and manage malicious participants. It utilized blockchain's tamper-proof properties to enhance security and privacy in distributed training. To address the privacy leakage risks, authors in [172] proposed a dualmodule system combining blockchain for secure data transmission and DL for detecting malicious vehicles using data secured by the blockchain module.

For intrusion detection, [173] proposes a collaborative learning framework integrating FL and blockchain to enhance security, reduce latency, and protect privacy in vehicular edge computing. Blockchain ensures tamper-proof training, while FL enables private collaborative learning. Expanding on this, [174] proposed a DL-based intrusion detection system using vehicle nodes for distributed training. It consists of a converter network for attack classification and blockchain for reliable distributed training by preventing unreliable updates and securing the integrity of the learning process.

3) Homomorphic Encryption: Studies in the literature integrated HE into in-vehicle monitoring and secure communication, enabling remote data processing while preserving user and vehicle privacy. Authors in [175] utilized HE for secure neural network classification on encrypted data to detect distracted drivers. Building on [176], it employs FHE schemes and transciphering for efficient encryption, even in resource-constrained automotive settings. HE-friendly activation functions enabled privacy-preserving inference. In the proposed framework a homomorphic computation server (HCS) handled encrypted calculations and maintained the separation of keys so that no single entity could decrypt the private data. However, the approach is limited by the timeintensive preprocessing and encryption during training.

To secure sensitive driver data during training, the authors in [124] proposed a privacy-preserving framework for driver drowsiness detection (PFTL-DDD) utilizing FL. This framework employs CKKS HE to protect data throughout the training process while transferring knowledge from a pretrained model to a FL framework, enhancing both accuracy and privacy. Similarly, study [177] introduced PRIV-DRIVE, which combines FL with Paillier HE (PHE) for driver fatigue detection. It encrypted model parameters using PHE and top-k selection, preserving data privacy while optimizing computation and communication. Both studies show HE enhances FL systems for secure and efficient in-vehicle monitoring.

4) Secure Multi-party Computation: In the literature SMPC is used in cooperative control, cruise control, vehicle classification, and speed advisory systems to protect data while enabling real-time decisions for autonomous driving safety. For cooperative control of CAVs, [178] proposed AutoMPC framework, which applied SMPC techniques like function secret sharing combined with a distributed oblivious random access memory to secure CAV data. It includes an adaptive proportional-derivative controller for latency and threats but lacks experimental validation for V2V communication reliability. Study in [179] proposed a privacypreserving control framework for CAVs in mixed traffic using affine masking to protect vehicle data. Built on data-enabled predictive leading cruise control (leveraging both Hankel and for non-parametric page matrix structures representation), it optimized traffic flow and improved fuel efficiency while ensuring rigorous privacy guarantees.

In another study, [180] proposed FedVPS, a framework leveraging SMPC to enhance privacy and security in FL for IoV. By integrating SMPC with differential privacy, FedVPS addressed non-IID data and model heterogeneity, using prototype-based aggregation to reduce communication costs while maintaining high accuracy. It outperforms baseline methods like FedAvg in vehicle classification tasks and ensures protection against privacy threats, including reconstruction and membership inference attacks. Similarly, a privacy-preserving consensus-based speed advisory system (CSAS) using SMPC was introduced [181], enabling vehicles to compute an optimal consensus speed without exposing private data. By securely sharing secret mappings like speedemission data, SMPC ensured privacy while allowing accurate decision-making in real-time.

5) Differential Privacy: In autonomous driving, in-vehicle monitoring and driving statistics collect a significant amount of sensitive information. Study in [182] evaluates the application of DP for in-vehicle monitoring and driving statistics, balancing privacy and service quality. The authors analyzed four DP mechanisms and proposed high utility (HUT) for batched queries. HUT enhanced data utility while preserving strong privacy guarantees.

TABLE IV SUMMARY OF EXISTING LITERATURE ADOPTING PPML IN AUTONOMOUS DRIVING AND SAFETY-CRITICAL APPLICATIONS

Application: Lane-change prediction: Data Application: Lane-change prediction: Data Application: Lane-change prediction: Data Application: Activity recognition: Data Type: Teamwork: English and PL. LSTM networks. Driver motions of the properties of the pro	Ref	Application and Data Type	Framework and Key Features	Performance evaluation
Dataset: StulyChen dataset Application: Steering angle prediction: Data Type: Challenge and StateForm dataset Application: Driving information on different rouses Parameters Feed Feed Feed Feed Feed Feed Feed Fee	1. Fed	erated Learning:		
Privacy Parameter Superior Privacy Processing Privacy Parameter Superior Privacy P	[118]	Type: Station camera images, Passenger location;	Clustering-based FL, LSTM networks, Driver	prediction accuracy; Efficiency: 7.6x higher
Patiests: SPIDD and Driver actions and steering, mirror checks, etc.)	[121]	Driver action videos/images; Dataset: AI City	Features: Edge device optimization, Video	Recognition rate (98.9%); Efficiency: Reduce
Dataset: NTH-DDD and VAWDD dataset Protection Prote	[123]	vehicle sensor data (steering, mirror checks, etc.);	Transfer learning, Ordered aggregation,	462% better accuracy; Efficiency : 37.46% lower
Continual learning, Peer-to-peer architecture, Challenge and StateFarm dataset Application: Steering angle prediction: Data Application: Steering angle prediction: Data Particular of The New York of State Presentation Type: In-whelice camera; Data Type: Data Set: Sully Chen dataset Application: CaV motion control: Data Type: Claid Wehicle control data; Dataset: Data Type: Intersection traffic data; Application: Steering angle prediction: Data Type: Particular of State States of Control and collision Application: CaV motion control: Data Type: Paramework: Fl-hased control; Features: New York of States of Control and collision Application: Stafe Braking: Data Type: Rear-end collision sensor data; Dataset: Generated dataset or Data States of Control data; Dataset: Generated dataset or Data States of Control data; Dataset: Generated dataset or Data States of Control data; Dataset: Generated dataset or Data States of Control data; Dataset: Generated dataset or Data States of Control data; Dataset: Generated dataset or Data States of Control data; Dataset: Gala Open Dataset: Application: Autonomous driving: Data Type: Vehicle trajectory data; Dataset: Control data Control dataset: Data States or Data S	[124]*	Driver fatigue indicators (camera signals);	CKKS encryption, Transfer learning, Privacy	Utility: High classification accuracy; Efficiency:
Dataset: Carno Safe Parking: Data Type: Investigation: Data Type: Investigation: Safe Parking: Data Type: Investigation: Carno	[126]	Driver action videos/images; Dataset : AI City	Continual learning, Peer-to-peer architecture,	High recognition accuracy; Efficiency: Low
Type: In-wehicle camera: Dataset: CarND self- driving car simulator by Udacity	[129]	Type : Driving information on different routes;	Distributed learning, Model aggregation, Privacy	prediction accuracy; Efficiency: 25% bandwidth
Vehicle control data; Dataset: BDD and DACT dataset Authonomous control	[130]	Type: In-vehicle camera; Dataset: CarND self-		accuracy; Efficiency: 62-250x lighter load on
avoidance; Data Type: Intersection traffic data; IoV integration, Unsignalized intersections, Dataset: N/A Dataset: Generated dataset Privacy: Data protection; Utility: 2.58% better accuracy for braking; Efficiency: 63% reduction in communication voerhead communication Autonomous braking communication Autonomous braking communication overhead communication Autonomous braking communication Autonomous braking communication Oxotehad Privacy: Data protection, Spatial-temporal security; Utility: High accuracy; Efficiency: Communication cost Privacy: Data protection, Spatial-temporal security; Utility: High accuracy; Efficiency: Communication cost Privacy: Data Type: Pramework: Features: Data Type: One-class SVM, Isolation forest, BiGAN integration integration integration integration integration Privacy: Privacy: Privacy: Raw data protection, Trajectory privacy; Utility: 98-99% accuracy; Efficiency: Privacy: Privacy: Privacy: Raw data protection, Trajectory privacy; Utility: 98-99% accuracy; Efficiency: Privacy: Note of Privacy: Model explainability, Trust verification; Utility: Privacy: Model explainability, Trust verification; Privacy: Privacy: Model explainability, Trust verification; Privacy: Privacy: Privacy: Privacy: Privacy: Model explainability, Privacy: Privacy: Model explainability, Privacy: Privacy: Model explainability, Privacy: Privacy: Model explainability, Privacy: Privacy: Privacy: Model explainability, Privacy: Privacy: Privacy: Privacy: Model explainability, Privacy: Data Individual Privacy: Privacy: Privacy: Privacy: Data Individual Privacy: Privacy: Privacy: Data Individual Privacy: Privacy: Privacy: Data Individu	[134]	Vehicle control data; Dataset : BDD and DACT	Incentive mechanism, Edge computing,	Superior velocity tracking; Efficiency: 40%
Application: Sale Friangis, Data Type: Application: Vehicle trajectories; Data Type: Vehicle trajectories; Data Type: Vehicle trajectories (Dataset: Gala Application: Anomaly in vehicle trajectories; Data Type: Vehicle trajectory data; Dataset: Application: Anomaly in vehicle trajectories; Data Type: Vehicle trajectory data; Dataset: Pramework: Fedared detection; Features: One-class SVM, Isolation forest, BiGAN integration Application: Trajectory prediction; Data Type: Vehicle trajectory data; Dataset: NTERACTION dataset Application: Trajectory prediction; Data Type: Vehicle trajectory data; Dataset: NGSIM dataset Proxy re-encryption	[135]	avoidance; Data Type: Intersection traffic data;	IoV integration, Unsignalized intersections,	Utility: Motion accuracy; Efficiency: 0.44%-
Spatio-temporal FL, s-FedWvg and t-FedWvg	[136]		Features: Collision avoidance, Efficient	accuracy for braking; Efficiency: 63% reduction
Data Type: Vehicle trajectory data; Dataset: INTERACTION dataset Negligible N	[139]	Vehicle trajectory data; Dataset: GAIA open	Spatio-temporal FL, s-FedWvg and t-FedWvg,	security; Utility: High accuracy; Efficiency:
Three-layer architecture, FAHE1 encryption, Vehicle trajectory data; Dataset: NGSIM dataset Vehicle trajectory data; Dataset: NGSIM dataset: Pramework: XAI-based FL; Features: Deep Privacy: Model explainability, Trust verification; Driving behavior and environment data; Dataset: Motion prediction dataset Pramework: XAI-based FL; Features: Deep Privacy: Model explainability, Trust verification; Peature contribution analysis Privacy: Model explainability, Trust verification; Utility: 95% accuracy; Efficiency: Real-time processing, high convergence search, Multi-source federation Privacy: Model explainability, Trust verification; Peature contribution analysis Privacy: Model explainability, Trust verification; Utility: 95% accuracy; Efficiency: Real-time processing, high convergence search, Multi-source federation Privacy: Model security, Design privacy; Utility: Superior accuracy; Efficiency: Better search efficiency. Privacy: Local data, Model personalization; Utility: 2x better than baseline; Efficiency: Resource optimization. Privacy: Application: Contextual trajectory prediction; Data Type: Data Type: Local multi-agent point cloud data; Dataset: Generated from CARLA simulation Pramework: Dynamic map fusion; Features: Obstact Type: Local multi-agent point cloud data; Dataset: Generated from CARLA simulation Privacy: Data Interprivacy: Utility: Privacy: Model personalization; Utility: 2x better than baseline; Efficiency: Resource optimization. Privacy: Model personalization; Utility: 2x better than baseline; Efficiency: Resource optimization. Privacy: Model personalization; Utility: 2x better than baseline; Efficiency: Resource optimization. Privacy: Model privacy; Utility: Efficiency: Real-time fusion. Privacy: Data locality; Utility: Comparable to centralized (68% mean average precision); Efficiency: Communication cost collaboration. Privacy: Data locality, Model privacy; Utility: Enhanced detection; Efficiency: Real-time processing	[141]	Data Type: Vehicle trajectory data; Dataset:	One-class SVM, Isolation forest, BiGAN	Utility: 98-99% accuracy; Efficiency: Parameter
Privacy: Model security, Design privacy; Utility: Synthetic dataset from CARLA simulation pataset: Generated from CARLA simulation pataset: Gisting Distributed image datasets; Dataset: KITTI Vision Benchmark 2D image dataset Distributed image dataset; Dataset: KITTI vision Benchmark 2D image dataset: CADC Application: Snow detection; Data Type: Snowy weather object detection data; Dataset: CADC attaset. Spatiaset: CADC attaset. Spatiaset. Spatiaset: CADC attaset. Spatiaset: CADC attaset. Spatiaset. Spatiaset. CADC attaset. Spatiaset. Spatiaset. CADC attaset. Spatiaset. Spatiaset. Spatiaset. Spatiaset. Spatiaset. Spatiaset. Spatiaset. Spatiaset. CADC attaset. Spatiaset. Spa	[142]*	Application: Trajectory prediction; Data Type: Vehicle trajectory data; Dataset: NGSIM dataset	Three-layer architecture, FAHE1 encryption,	Utility: Over 99% accuracy; Efficiency: High
Type: Trajectory datasets; Dataset: ETH and UCY Dataset Application: Trajectory prediction; Data Type: Driver behavior and trajectory data; Dataset: Synthetic dataset from CARLA Application: Contextual trajectory prediction; Data Type: Data Type: Local multi-agent point cloud data; Dataset: Generated from CARLA simulation Application: Object detection; Data Type: Distributed image datasets; Dataset: KITTI Vision Benchmark 2D image dataset Application: Snow detection; Data Type: Snowy weather object detection data; Dataset: CADC dataset: Weather object detection data; Dataset: CADC dataset: Willing, Safety enhancement Automatic model design, Relation-sequence search, Multi-source federation Application: Trajectory prediction; Data Type: Framework: Personalized FL; Features: Robust learning, User distribution adaptation Framework: Personalized FL; Features: Robust learning, Framework: Dynamic map fusion; Features: Three-stage fusion, Features model tuning, Knowledge distillation Framework: Personalized FL; Features: Robust Utility: 2x better than baseline; Efficiency: Resource optimization Privacy: Map privacy; Utility: High fidelity; Efficiency: Real-time fusion Framework: FL-YOLO; Features: Distributed learning, Model aggregation, Vehicle collaboration Framework: FL-YOLO-CNN framework; Fficiency: Communication cost Privacy: Data locality, Wodel privacy; Utility: Enhanced detection; Efficiency: Real-time processing	[143]	Driving behavior and environment data; Dataset :	reinforcement learning, Trust computation,	Utility: 95% accuracy; Efficiency: Real-time
Privacy: Map privacy; Utility: Comparable to centralized ficency: Resource optimization	[144]	Type: Trajectory datasets; Dataset: ETH and	Automatic model design, Relation-sequence	Superior accuracy; Efficiency: Better search
Three-stage fusion, Feature model tuning, Knowledge distillation Privacy: Map privacy; Utility: High Indenty; Efficiency: Real-time fusion	[145]	Driver behavior and trajectory data; Dataset :		Utility : 2x better than baseline; Efficiency :
[153] Distributed image datasets; Dataset : KTTTI Vision Benchmark 2D image dataset Application: Snow detection; Data Type : Snowy weather object detection data; Dataset : CADC dataset Framework: FL-YOLO-CNN framework; weather object detection data; Dataset : CADC dataset Framework: FL-YOLO-CNN framework; privacy: Data locality, Model privacy; Utility : Enhanced detection; Efficiency : Real-time processing	[146]	Data Type: Local multi-agent point cloud data;	Three-stage fusion, Feature model tuning,	
[154] weather object detection data; Dataset : CADC dataset Features: YOLO adaptation, Weather-specific training, Safety enhancement Enhanced detection; Efficiency: Real-time processing	[153]	Distributed image datasets; Dataset: KITTI	learning, Model aggregation, Vehicle	centralized (68% mean average precision);
	[154]	weather object detection data; Dataset : CADC	Features: YOLO adaptation, Weather-specific	Enhanced detection; Efficiency: Real-time

(Continued)

[155]	Application: Road user classification; Data Type: LiDAR data; Dataset: nuScenes dataset	Framework : Decentralized FL; Features : V2X networks, Road user categorization	Privacy : User privacy; Utility : Outperforms self-learning methods; Efficiency : Optimize Network
[156]	Application : 3D Pothole detection; Data Type : 3D road surface data; Dataset : Generated dataset	Framework: 3Pod framework; Features: Depth extraction, Risk scoring, Crowd voting for maintenance	Privacy : Spatial privacy, Crowd security; Utility High vison-based pothole detection; Efficiency : Lightweight framework
[157]	Application: Road damage detection; Data Type: Road damage data; Dataset: Generated	Framework: FedRD framework; Features: Adaptive learning, Warning system, Privacy preservation	Privacy : Data protection, User privacy; Utility : Warning accuracy; Efficiency : Reduced computation and communication costs by 3/4
[158]	Application: Surface classification; Data Type: Road surface image datasets; Dataset: RSCD dataset	Framework: FedRSC framework; Features: Multi-label learning, Road classification, Federated analysis	Privacy : Data protection, Classification security; Utility : Classification accuracy; Efficiency : Low communication cost
[159]	Application : Road damage; Data Type : Road damage images; Dataset : JRDD and MRDD dataset	Framework : FL-based detection; Features : Global model training, Damage classification, Multi-region learning	Privacy: Regional privacy, Kodel integrity; Utility: 1.33%–163% improvement in accuracy; Efficiency: Global optimization
[160]	Application: License plate recognition; Data Type: License plate images; Dataset: Fujian traffic police dataset	Framework: FL-based recognition; Features: License plate recognition, Distributed learning, Edge computing	Privacy: Data protection, Network security; Utility: High accuracy; Efficiency: Low latency
[161]	Application: Traffic sign recognition; Data Type: Traffic sign images; Dataset: BelgiumTS dataset	Framework: Spike neural networks; Features: Efficient training, Resource optimization, Sign recognition	Privacy : Model security; Utility : Superior accuracy; Efficiency : High energy efficiency and noise resistance
[162]	Application: Resource optimization; Data Type: Multi-modal perception datasets; Dataset: Generated using CarlaFLCAV	Framework: Federated perception; Features: Multi-modal fusion, Design verification, Perception models	Privacy : Model security, Sensor privacy; Utility : High perception accuracy; Efficiency : Low network resources
[164]	Application: Resource and performance optimization; Data Type : Vehicular edge image data; Dataset : MNIST and BelgiumTSC dataset	Framework: Selective aggregation; Features: Model selection, Resource optimization, Edge deployment	Privacy: Model security, Vehicle privacy; Utility: Better aggregation accuracy; Efficiency: Better resource and aggregation efficiency
[165]	Application: Resource optimization; Data Type: Thermal image and driving records; Dataset: FLIR and BDD100K dataset	Framework: AF-DNDF framework; Features: Asynchronous learning, Deep neural decision forests, Distributed training	Privacy : Training privacy; Utility : Average 5% improvement than centralized; Efficiency : Lower training time (60%) and bandwidth cost (80%)
2. Bloc	ckchain-based PPML:		
[166]	Application: Route optimization and emergency response; Data Type : Autonomous Vehicle data; Dataset : Generated dataset	Framework: BCL framework; Features: Distributed training, Collective intelligence, Blockchain verification	Privacy : Data locality; Utility : Lowest position error; Efficiency : Reduced transmission
[167]	Application: Autonomous decision making; Data Type: CAV driving behavior and traffic data; Dataset: Generated dataset	Framework: Hybrid strategy; Features: Rule extraction, Vehicular blockchain validation, DRL + Expert System, BFT-DPoS consensus	Privacy: Knowledge privacy, Decision transparency; Utility: High decision accuracy; Efficiency: Adaptive learning
[168]	Application : Vehicle positioning; Data Type : GPS positioning error data; Dataset : Generated	Framework : Blockchain framework; Features : Edge server computation, GPS error evolution, Smart contract automation, PBFT and BFT-DPoS	Privacy : Location privacy, Cooperative security; Utility : Low positioning error; Efficiency : Real-time correction
[169]*	Application: Misbehavior detection; Data Type: Vehicular network data; Dataset: Generated	Framework: Blockchain-FL; Features: Trustworthy updates, Secure aggregation, Distributed consensus, PBFT consensus	Privacy : Model security; Utility : 97% accuracy, better throughput; Efficiency : low latency and energy consumption
[170]*	Application: Autonomous driving; Data Type: Autonomous driving data; Dataset: Rancho Palos Verdes and San Pedro California dataset	Framework: Privacy-preserved FL; Features: DGHV algorithm, Reputation-based incentives, Zero-knowledge proofs, PoA, PoW consensus	Privacy : Identity and message confidentiality; Utility : Improve accuracy by 5.55%, 99% privacy; Efficiency : Reduce 73.7 % training loss
[171]*	Application: Misbehavior detection; Data Type: VANET data; Dataset: VeReMi dataset	Framework: FL-Blockchain; Features: Gaussian mechanism, Edge coordination, Differential privacy, Enhanced delegated PoA	Privacy : DP guarantee, data privacy; Utility : High detection accuracy; Efficiency : Resource optimization, Low overhead
[172]*	Application: Privacy leakage prevention; Data Type : IoV intrusion detection data; Dataset : IoT-Botnet and ToN-IoT dataset	Framework: P2SFIoV framework; Features: Multi-layer security, Authentication mechanism, Secure communication, ePoW consensus	Privacy : Data protection; Utility : High detection rate; Efficiency : Resource optimization, Scalable
[173]*	Application: Intrusion detection; Data Type: Vehicular network intrusion data; Dataset: KDDCup99 dataset	Framework: Collaborative IDS; Features: Distributed training, Edge offloading, Secure aggregation, PoW and PoA consensus	Privacy : Storage and sharing privacy; Utility : High detection rate; Efficiency : Low resource utilization
[174]*	Application: Intrusion detection; Data Type : Smart transportation system attack data; Dataset : Car-Hacking, TON-IoT dataset	Framework: FED-IDS; Features: Context- aware transformer, Blockchain management, Distributed training, dBFT	Privacy : Blockchain update privacy; Utility : High attack detection; Efficiency : Distributed processing, Reliable training
3. Hon	nomorphic encryption:		
[175]	Application: Distracted-driver detection; Data Type : Driver behavior and vehicle data; Dataset : State Farm dataset	Framework : HE framework; Features : Homomorphic computation server, Privacy-aware protocols, Secure computation	Privacy: Data confidentiality, Service privacy; Utility: 86.2% classification accuracy; Efficiency: long classification time
	Application: Driver monitoring; Data Type:	Framework: Federated transfer learning;	Privacy: Behavior privacy, CKKS-based

[177]*	Application: Driver fatigue detection; Data Type: Driver fatigue detection data; Dataset: UTA-RLDD dataset	Framework: HE-based FL; Features: Paillier encryption, Top-k selection, Encrypted model updates, PoC implementation	Privacy: Enhanced security, Driver privacy; Utility: 76% accuracy; Efficiency: Reduced 60- 96% computation time and 95% traffic
4. Seci	ure Multi-party Computation:		
[179]	Application: CAV control; Data Type: Mixed traffic flow data; Dataset: NGSIM dataset	Framework: Predictive and cooperative CAV control; Features: Affine masking, Privacy-preserved optimization, State concealment	Privacy: State privacy, Input confidentiality; Utility: Safe/optimal CAV control; Efficiency: Balanced overhead, Low computation (<30ms)
[180]*	Application: Vehicle classification and attack prevention; Data Type: Non-IID IoV data; Dataset: BIT-Vehicle dataset	Framework: FedVPS; Features: SMPC protection, Heterogeneous FL, Edge-cloud architecture	Privacy: Terminal privacy, Model privacy; Utility: High prediction accuracy; Efficiency: Improved communication efficiency
[181]	Application: Speed advisory: Data Type: Vehicle speed and advisory data; Dataset: Generated from SUMO	Framework: MPC-CSAS; Features: Real-time MPC, Privacy preservation, One-iteration convergence	Privacy : Data privacy; Utility : Optimal speed advisory; Efficiency : Single iteration convergence, Lightweight, Dynamic network
4. Diff	erential Privacy:		
[182]	Application: In-vehicle monitoring and driving statistics; Data Type : IoV trajectory and communication data; Dataset : BROOK dataset	Framework : Optimized distributed DP; Features : Continual observation, Route-level privacy, Hybrid noise scheme	Privacy: Real-time privacy, Route anonymity; Utility: Improved data utility on frequently-used batched queries; Efficiency: Resource optimization, Reduce information loss by 95.69%
[183]	Application : Trajectory publishing; Data Type : Passenger trajectory data; Dataset : Montreal bus and metro transit networks dataset	Framework : SafePath framework; Features : Passenger path anonymization, Adaptive privacy budget, Path clustering	Privacy: Trajectory privacy, Data utility preservation; Utility: Comparable data utility; Efficiency: Improved runtime and scalability
[184]	Application: Enhanced trajectory partitioning; Data Type: Vehicle trajectory data; Dataset: Generated from SUMO	Framework: DP-guaranteed scheme; Features: Trajectory community detection, Privacy-aware clustering, Background knowledge protection	Privacy : community privacy; Utility : Reduced info loss, Outperforms k-anonymity; Efficiency : Improved information loss and efficiency
[185]	Application: Vehicle trajectory; Data Type: Vehicle trajectory data; Dataset: Geolife, ShangHai dataset	Framework: Real-time privacy; Features: Dynamic privacy budget, Ensemble Kalman filter, Trajectory perturbation	Privacy: Charging privacy, Location anonymity; Utility: Data availability, high prediction accuracy; Efficiency: High budget allocation
[186]	Application : Intrusion detection; Data Type : VANET intrusion detection data; Dataset : NSL-KDD	Framework: PML-CIDS framework; Features: ADMM optimization, Dynamic DP, Dual variable perturbation	Privacy : Collaborative privacy; Utility : High intrusion detection rate; Efficiency : Scalable implementation
[187]	Application: Intrusion detection; Data Type: VANET intrusion detection data; Dataset: NSL- KDD dataset	Framework : SP-CIDS; Features : DML with ADMM, Ensemble classifiers, DP integration	Privacy : Multi-level security, Training privacy; Utility : 96.94% accuracy; Efficiency : Enhanced storage and computation
	Application: Intrusion detection; Data Type: Inter-vehicle network data; Dataset: VeReMi Extension dataset	Framework: DPFL-F2IDS; Features: LSTM-based detection, Member inference defense, Differential privacy	Privacy: Model privacy; Utility: 0.97-0.95 F1- score; high threat detection, Efficiency: Distributed processing, high computation
[189]*	Application : Secure communication; Data Type : Trajectory communication data; Dataset : N/A	Framework: LDP-IOTA; Features: IOTA ledger, LDP, Distributed architecture	Privacy : Vehicle privacy; Utility : System reliability; Efficiency : Resource optimization

Real-time reporting in IoV systems raises privacy concerns, as vehicular trajectory data can reveal sensitive behavioral patterns. SafePath [183] mitigated this risk using DP by constructing a noisy prefix tree for secure trajectory publication. Another study, [184] integrated exponential DP with trajectory partitioning and clustering to enhance efficiency and data utility while reducing information loss. Similarly, [185] combines a dynamic sampling strategy with a Kalman filter, adding Laplace noise to balance data availability and privacy of vehicle trajectory.

Various CAV and traffic management systems use intrusion detection systems (IDS) to enhance security and potential threats. However, data breaches can compromise IDS training to avoid detection or for other malicious purposes. To address this, [186] proposed a DP-based ML IDS for VANETs, using alternate-directional multipliers and dual variable perturbation to balance security and privacy. Similarly, [187] introduced a secure and private IDS, which is a distributed ML IDS integrating DP with the alternating direction method of multipliers (ADMM) to protect V2V communication. Expanding on this, [188] developed a differentially private FL framework (DPFL-F2IDS) to prevent membership inference

attacks in IDS while optimizing the utility-privacy trade-off.

Few studies integrated blockchain with DP to enhance privacy and trust of IoV systems. Authors in [171] proposed a blockchain-based FL scheme with DP for misbehavior detection in VANETs. It leveraged differential privacy with the Gaussian mechanism to provide strict privacy protection for the model on the blockchain, ensuring data security and privacy while coordinating multiple distributed edge devices. Another study integrated LDP with the IOTA distributed ledger for privacy-preserving framework in IoV [189]. It leveraged privacy guarantees of LDP and distributed ledger technology of IOTA to achieve scalability, immutability, and quantum resistance in large-scale vehicular networks.

C. Communication Infrastructure and Smart Services:

Modern IoV ecosystems rely on robust communication infrastructures for smart services like EV charging, predictive maintenance, smart parking and context-aware infotainment. Ensuring data privacy while maintaining performance is a key challenge. PPML techniques can enhance security and privacy in heterogeneous networks protecting sensitive data in usercentric services. This section reviews PPML techniques that contributed to enhancing privacy, security, and functionality

within IoV communication infrastructures and smart services.

Table. V. provides technical summary of studies adopting PPML in IoV communication infrastructure and smart services.

1) Federated Learning: Vehicular cyber-physical systems (VCPS) use V2X communication to integrate vehicles, infrastructure, and traffic management. These systems contain sensitive vehicular and user data, which are vulnerable to leakage and unauthorized access. FL enhances privacy by enabling distributed learning on local data points while preserving heterogeneity and minimizing data exposure [46]. Authors in [190] designed a FL model to prevent data leakage within VCPS. Study in [191] proposed OES-Fed that improved anomaly detection in vehicular networks by noise filtration. It enabled systems to identify and mitigate abnormal data inputs without transferring vehicle-specific data. Another adaptation is the application of extreme value theory (EVT) and personalized FL in [192], which address heterogeneous data distributions among vehicles by modeling rare, anomalous events while preserving data privacy across a highly variable vehicular network. The study in [193] integrates deep Q-network (DQN) with FL to significantly reduce latency in vehicular data sharing, enabling secure and efficient communication within vehicular networks while preserving user privacy. Furthermore, the resilience of FL to adversarial attacks in VCPS is enhanced through its integration with other PPML techniques such as DP [194] and blockchain-leveraged methods [195]. These frameworks further enhanced VCPS security and reliability, while preserving vehicle data privacy.

The smart parking control system is essential for managing urban parking, which mostly incorporates a request center and an assignment center that uses sensors for availability checks and reservations. Most smart parking providers use third party cloud storage to centralize data which raises privacy concerns. To address this, [196] proposed a FL framework for real-time parking predictions, allowing vehicles to forecast availability without sharing sensitive data and introducing an incentive mechanism to enhance participation and prediction accuracy. With the rise of EVs, parking areas now integrate charging stations (CSs) managed through centralized systems, which optimize charging but face risks like system crashes and privacy breaches [47]. FL frameworks have been studied to enhance privacy by grouping EVs and using sub-aggregators to manage local data processing. Studies have integrated FL with blockchain and ML techniques, such as random forests and CNNs, for power load prediction [197], while clusteringbased approaches have reduced communication costs and prediction bias [198]. However, the diverse characteristics of EV and CS remain a challenge for collaborative learning. A [199] cross-platform FLframework in combined recommendation models with encryption techniques like hash and RSA to balance privacy and real-time prediction accuracy. Economic-driven FL model proposed in [200], further optimize multi-agent charging scenarios, where a multiprincipal one-agent (MPOA) model transforms CS utility

maximization problem into a decentralized non-cooperative energy optimization framework, ensuring privacy-aware resource sharing. Furthermore, unpredictable consumption patterns cause energy demand fluctuations in EV systems, challenging traditional FL models. To address this, [201] enhanced FedAvg with a probabilistic algorithm for better adaptability, while operators and aggregators balance cost, efficiency, and resource optimization.

Route planning is an intelligent service which requires consideration of dynamic obstacles and external conditions in complex road networks, including road layouts and traffic flows. Centralized systems often face latency issues, reducing decision-making efficiency. To address this, [202] proposed a FL-based decentralized approach using fog nodes and RSUs, reducing memory usage, latency, and communication overhead. In dynamic traffic environments, modeling roads as time-dependent graphs further improved route optimization. Clustering technique in [203], balanced computational loads across edge nodes, enhancing data processing efficiency. Using A* algorithm on time-dependent graphs enabled accurate route selection while preserving privacy by minimizing cloud data transmission. Hierarchical clustering further optimized traffic predictions without data sharing.

In addition to route planning, FL is increasingly applied in smart user applications like travel time estimation and destination prediction, balancing privacy with model accuracy. In travel time estimation, a global model is trained using data from all participants, while personalized models are fine-tuned for individual driving patterns, ensuring privacy [204]. FL has also been used for cross-area travel time estimation, where localized models are trained in different regions and combined through FL to preserve privacy across geographic boundaries [205]. Additionally, FL has been applied to destination prediction tasks, providing precise location services without exposing sensitive user data [97]. The framework improved localization in areas with poor GPS signals by using unmanned aerial vehicles (UAVs) as aerial anchors [206]. FL techniques have also been used to aggregate models from edge devices, optimizing localized path predictions and reducing localization errors [207].

As IoV services expand, the rise in connected vehicles, devices, and infrastructure increases data transmission, posing challenges in communication efficiency, energy use, and privacy. FL addresses these issues by selecting clients and servers efficiently during training [208], [209], which significantly improved resource management and system responsiveness [210], [211]. By training models locally and aggregating only learned parameters, FL reduces large-scale data transmission in resource-limited vehicular networks. Author in [212] proposed a CNN-based FL framework for 6G IoV environments aiming to enhance model quality through hierarchical aggregation at edge and cloud levels. The approach considered factors such as RSU proximity and vehicle density. In [86], remote sensing image analysis focused on vehicle target recognition, leveraging data from diverse environments. FL was utilized to overcome the

limitations of single-node data processing without compromising the privacy of sensitive geospatial information. Furthermore, techniques such as EVT and Lyapunov optimization were employed to optimize FL frameworks, enabling better handling of anomalous events and dynamic power allocation [213].

2) Blockchain-based PPML: There are hybrid frameworks combining blockchain, FL, and DP to enhance data security and resilience in VCPS [194]. Study in [214] proposed authentication scheme between vehicles and RSUs utilizing blockchain. Using on-chain hashing, off-chain integrity certificate schemes. cryptographic algorithms, and authentication, the system ensured anonymous service requests, two-way authentication, and privacy preservation. Dynamic pricing in the IoV ecosystem requires real-time data handling with transparency and fairness. A hybrid approach in [215] integrates blockchain for secure transactions between vehicle owners and regulatory bodies and DL for traffic prediction, ensuring data reliability and payment transparency. To further optimize resources and transactions, a privacypreserving energy trading scheme in [216] uses blockchain and zero-knowledge proofs, ensuring confidentiality in energy transaction between EVs and the power grid. Decentralized identifiers anonymized participants, while smart contracts enforced fair pricing without intermediaries.

Incorporating software-defined networking (SDN) and blockchain into IoV applications enhances privacy and security in distributed environments. Authors in [217] proposed a 5G-enabled fog computing paradigm where RSUs act as SDN controllers, managing blockchain operations and secure channel selection. This decentralized approach reduces reliance on central servers and implements reputation-scoring mechanisms for security. Similarly, [218] presented dual-layered SDN-controlled vehicle edge computing (VEC) framework integrating blockchain for secure network topology sharing. By using an enhanced PBFT algorithm, it improved system throughput, reduced latency, and ensured data integrity in SDN operations.

In IoV, real-time data transmission among RSUs, vehicles, users, and central management systems is crucial for ML applications. Ensuring trust and privacy in these communications is challenging. To address this, authors have proposed integrating blockchain with FL. In [219], a blockchain-empowered FL framework enabled distributed intelligence while preserving privacy. The blockchain architecture ensured secure, traceable interactions among the untrusted entities. To further enhance the efficiency of knowledge sharing, authors in [220] proposed a hierarchical blockchain framework combined with a layered FL approach. It consisted of multiple leader and player setup, where vehicles function as individual FL nodes. To mitigate blockchain overhead, a lightweight Proof-of-Knowledge (PoK) consensus mechanism was introduced, optimizing resource utilization while maintaining data privacy.

Privacy concerns in vehicular social networks are studied in [221] through a secure SVM classifier training system based

on blockchain and cryptographic techniques. It eliminated third-party intermediaries through smart contracts, which leveraged privacy-preserving protocols to ensure data confidentiality and utilized blockchain's decentralization for security. A blockchain-powered autonomous FL system for vehicular communication was proposed in [222], optimizing parameters like block size and arrival rate to enhance efficiency. Additionally, a blockchain-based FL solution for emergency message dissemination was introduced in [195]. It leverages Proof-of-FL (PoFL) consensus to mitigate broadcast storms and low packet reception, and a Stackelberg game-based model to incentivize participation in model training to achieve improved accuracy.

3) Homomorphic Encryption: Existing literature has studied the use of HE to address privacy concerns in various IoV applications including charging location privacy, vehicle tracking prevention, and secure interactions within ridesharing platforms. Studies such as [223] proposed a privacypreserving distributed matching algorithm for EV charging. It leveraged the Paillier cryptosystem to secure location data and employed bichromatic mutual nearest neighbor (BMNN) computation to identify and connect with nearby suppliers through local communication without exposing sensitive information. Similarly, [224] introduced PADP, a privacypreserving data aggregation and dynamic pricing scheme for V2G networks. It used HE to protect power consumption data while enabling real-time aggregation for dynamic pricing across regions. Additionally, [225] developed a blockchainbased federated DL framework for EV charging demand prediction, using CKKS HE to ensure privacy during model training while maintaining prediction accuracy.

Several studies in the IoV ecosystem highlight the utility of HE for various services. For ride-sharing, [226] proposed a framework using PHE to protect sensitive data. It ensured privacy-aware ridesharing and routing by protecting sensitive data, such as origins and destinations, during communications. To prevent vehicle tracking, [227] introduced a privacy risk assessment model that evaluated risks associated with toll transponders. The model utilized lattice-based FHE establishing a foundation for post-quantum cryptographic solutions in IoV scenarios. Study in [228], combined HE with blockchain for intelligent transportation, employing partially hashing HE and decryption (PHHE/D) for local data encryption at fog nodes. The secure, cost-optimal workload assignment (SCWA) algorithm ensured efficient processing, while the blockchain enhanced security and operational efficiency in the IoV network.

To enhance data security and confidentiality in V2X communication, several studies have integrated HE in the network. In [229], a privacy-aware intelligent forwarding solution, PABRFD, is introduced for named data networking(NDN)-VANETs. It integrated HE with an enhanced Bayesian receiver forwarding decision (BRFD) mechanism to enable secure and private vehicle-to-base (V2B) data exchange. Additionally, authors have studied combining HE with blockchain for further security enhancements. In

[104], a decentralized privacy-preserving DL (DPDL) model for VANETs integrated FHE and blockchain to enable secure data exchanges among vehicles and edge nodes. It ensured data privacy and mitigated threats like model extraction. Another study [230], proposed a privacy-preserving computing scheme for VANETs that combines PHE with directed acyclic graph (DAG) blockchain instead of traditional blockchain. This approach ensured data privacy, enabled computations on encrypted data, and improved performance through parallel processing.

4) Secure Multi-party Computation: Studies utilizing SMPC focused on enhancing the efficiency and security of services like secure data sharing, communication, and personalized recommendation systems. Some integrated hybrid frameworks combining multiple techniques. For instance, the study [231] introduced an AI-powered blockchain framework that combined SMPC with advanced cryptographic techniques to protect the privacy of vehicular data in ML applications. It decentralized key operations, used AI-driven smart contracts to prevent data exposure, and leveraged blockchain's tamper-proof ledger for secure, scalable data exchange. Another study in [232] proposed a game-theoretic SMPC framework for privacy-preserving data sharing in the IoV. The framework also decentralized data collection through distributed servers and utilized spatiotemporal maps to ensure privacy while maintaining utility. It incorporated Stackelberg game theory to optimize parameters such as payment and data-sharing frequency, and leveraged blockchain for transparent, reliable smart contracts.

To validate mobility data on the IoV ecosystem, the authors in [233] introduced BELIEVE framework. It is also a blockchain-enabled framework that utilizes SMPC to ensure privacy-preserving real-time validation. It employed a privacy-by-design approach utilizing encrypted distance-based computations for mobility data validation through a Proof-of-Presence (PoP) consensus mechanism. It is also supported by a permission blockchain and interplanetary file system (IPFS) for immutable storage, and an adaptive sampling strategy to minimize resource usage while protecting user privacy. Recommendation systems in in-vehicle infotainment (IVI) provide personalized content based on user preferences and behaviors. These systems rely on sensitive data, such as personal preferences, behaviors, and location, to tailor recommendations. The study [234] proposed a privacypreserving multi-party collaborative filtering system for IVI recommendations, using SMPC and the Paillier cryptosystem. It employed a symmetric balanced incomplete block design (SBIBD) for efficient aggregation in dynamic user groups which ensured sensitive data remains encrypted, while enabling accurate and timely location-based recommendations in vehicular networks.

5) Differential Privacy: In CAVs, smart charging and parking systems handle a significant amount of sensitive data, making them vulnerable to attacks and privacy breaches. Authors in [235] proposed differential privacy mechanism for

protecting sensitive EV charging data in V2G networks, using sampling intervals and sliding windows. Another study [236] proposed a privacy-preserving mechanism for EVs querying CSs. The authors propose approximate geoindistinguishability (AGeoI), which adapts geoindistinguishability vehicular for applications. This mechanism provides two-fold privacy protection safeguarding individual query locations and protecting against trajectory tracing in an online setting, all while maintaining high quality of service (QoS) for EVs.

Many CSs are located in or integrated with parking facilities, and while smart parking systems rely on continuous data sharing for features like occupancy monitoring and personalized recommendations, they inherently pose risks to user privacy. Study [237] proposed a privacy-preserving charging infrastructure system using ECC for mutual authentication and Laplace-distributed noise for LDP. This approach ensures data is perturbed before transmission, eliminating third-party anonymizers while preserving utility for recommendations. Extending this to broader vehicleinfrastructure ecosystems, study [238] addressed communication security between roadside infrastructure (CSs, traffic sensors, edge computation, etc.) and vehicles by integrating FL with LDP in VANETs. Here, vehicle data is perturbed locally before being shared with external infrastructures, enabling collaborative model training with a central server while mitigating gradient leakage and inference attack. To further defend against adversarial attacks exploiting vehicle speed and location, authors in [194] propose a privacypreserving VCPS. The system integrates FL with DP using the Laplace mechanism and applies layer-wise relevance propagation (LRP) to regulate perturbation values.

V. CHALLENGES AND FUTURE DIRECTION

PPML faces several challenges that hinder its wide-scale adoption. These challenges stem from inherent trade-offs between privacy, computational efficiency, and model performance, as well as technical complexities in implementing robust privacy mechanisms. When applied to the IoV, these issues are further compounded by the unique characteristics of IoV systems, such as dynamic network environments, heterogeneous data sources, and stringent latency requirements. This section reviews the challenges in PPML techniques and its adoption in the IoV ecosystem.

A. PPML Techniques

The key challenge in designing an optimal PPML solution lies in addressing the trade-off between different performance benchmarks. Current PPML approaches often compromise either system efficiency or utility (model performance) to achieve a desired level of privacy. Efficiency in traditional ML systems typically involves enhancing training or inference processes, especially for DNN architectures. However, in PPML systems, efficiency challenges manifest as communication efficiency, requiring minimal interactions and

$TABLE\ V$ Summary of existing literature adopting PPML in IoV communication infrastructure and smart services

Ref.	Application and Data Type	Framework and Key Features	Performance evaluation
1. Fed	lerated Learning:		·
[190]	Application: Data leakage prevention in VCPS; Data Type : Vehicular sensor and communication data; Dataset : 20 Newsgroups dataset	Framework : FL framework; Features : Data privacy preservation, Edge computing, Resource allocation	Privacy: Data locality; Utility: High accuracy and data leakage detection; Efficiency: Better computing utilization
[191]	Application: Outlier detection; Data Type: Noisy vehicular network data; Dataset: MNIST, CIFAR 10, BAIDU vehicle classification dataset	Framework: OES-Fed framework; Features: Data filtering, Quality assessment, Adaptive learning	Privacy: Data protection, Model integrity; Utility: Enhanced accuracy; Efficiency: Reduced noise
[192]	Application: Anomalous event detections; Data Type: Extreme event data in vehicle networks; Dataset: Simulation data	Framework : Personalized FL; Features : Event modeling, Personal adaptation, Network optimization	Privacy : Personal data, Event privacy; Utility : Event detection; Efficiency : Resource utilization, Reduced latency
[193]	Application: Collaborative secure data sharing; Data Type: Vehicular status data; Dataset: Generated dataset	Framework: FL-empowered framework; Features: Collaborative sharing, Edge computing, Resource management	Privacy: Data security; Utility: High accuracy; Efficiency: Low latency
[196]	Application: Parking space estimation; Data Type: Parking occupancy data; Dataset: Birmingham parking dataset	Framework: FedParking; Features: Edge assistance, Parked vehicle sensing, Real-time updates	Privacy : Location privacy; Utility : Accurate estimation; Efficiency : Resource and capacity optimization
[197]	Application: Power load prediction; Data Type: EV charging load time-series data; Dataset: Hangzhou charging station dataset	Framework: FL-based forecasting; Features: Load prediction, Federal architecture, Resource optimization	Privacy : Data protection; Utility : Prediction accuracy (below 3% loss); Efficiency : Resource utilization
[198]	Application: Energy prediction; Data Type: Charging station energy demand data; Dataset: Dundee city, UK charging stations dataset	Framework: Energy management in EV networks; Features: Demand forecasting, Network optimization, Federated training	Privacy: Energy privacy; Utility: Prediction accuracy (lowest RMSE 5.76%); Efficiency: Reduced communication overhead
[199]	Application: Station recommendation; Data Type: EV charging station usage patterns; Dataset: Generated dataset	Framework : FL framework; Features : Feature factorization, Entity alignment, Secure training	Privacy: Data locality; Utility: 6% AUC improvement, better convergence; Efficiency: Improved generalization ability and efficiency
[200]*	Application: Optimized multi-agent charging; Data Type: Charging station energy data; Dataset: Dundee city stations dataset	Framework: Contract theory FL; Features: Energy demand prediction, MPOA contracts, Profit maximization	Privacy: CS privacy, Vehicle privacy; Utility: outperform other economic models by 48%-36%; Efficiency: 88.9% lower communication
[201]	Application : Energy prediction; Data Type : EV energy demand and driving range data; Dataset : Generated dataset	Framework : Probabilistic FL; Features : Range prediction, Fleet learning, Uncertainty modeling	Privacy : Data protection; Utility : High probabilistic prediction; Efficiency : Optimum utilization of battery
[202]	Application: Decentralized route planning; Data Type: Real-time traffic and routing data; Dataset: Simulation data of mid-sized US city	Framework : Time-dependent FL; Features : Private fog networks, Online learning, Real-time processing	Privacy: Network privacy, Route privacy; Utility: Route optimization; Efficiency: Low latency and communication overhead
[203]	Application: Route selection, Traffic prediction; Data Type : Traffic flow and route planning data; Dataset : Simulation data from PeMS	Framework: Multi-task FL; Features: Hierarchical clustering, Route optimization, Time-dependent graphs	Privacy : Data protection, Task privacy; Utility : Improved accuracy; Efficiency : Task and route selection optimization
[204]	Application: Travel time estimation; Data Type: Travel time data (trajectories); Dataset: Didi Chengdu and Xi'an dataset	Framework: GOFTTE Framework; Features: Online generative model, Fine-tuned personalization, Client-side training	Privacy: Data locality, Model privacy; Utility: High accuracy (8-13% higher) compared to baseline; Efficiency: Reduced communication
[205]	Application: Cross-area travel trajectory; Data Type : Multi-region trajectory data; Dataset : Didi Chengdu and Xi'an dataset	Framework: Cross-area FL; Features: Uncertainty estimation, Bayesian deep learning, Monte-Carlo dropout	Privacy: Area protection, Trajectory privacy; Utility: Superior to baselines; Efficiency: Crossarea optimization
[209]	Application: Efficient client-server selection; Data Type: V2X messages; Dataset: MNIST, CIFAR-10 and SVHN dataset	Framework: V2X-boosted FL; Features: Contextual selection, V2X message fusion, Topology prediction	Privacy: Data protection; Utility: Enhanced contextual accuracy; Efficiency: Low convergence time
[210]	Application : Resource Management; Data Type : 6G-V2X data; Dataset : Simulation data	Framework: 6G-V2X Framework; Features: Computation offloading, Edge computing, Resource optimization	Privacy: Data security, Computation privacy; Utility: Resource efficiency; Efficiency: Reduced latency and computation cost
[211]	Application: Resource Management; Data Type: Vehicle-to-Vehicle (V2V) communication data; Dataset: Simulation data	Framework: Fed-MARL; Features: D3QN implementation, CSI optimization, Queue management	Privacy: Agent privacy; Utility: Efficient resource allocation; Efficiency: Network optimization, reduced delay
[212]	Application: Model quality enhancement; Data Type: IoV image data (traffic sign image); Dataset; BelgiumTSC dataset	Framework: Two-layer FL in 6G-IOV; Features: Heterogeneous aggregation, Model optimization, 6G integration	Privacy : Layer security; Utility : 96% average accuracy; Efficiency : Reduced communication overhead
[213]	Application: Power allocation, Anomalous event detection; Data Type: Vehicular communication data; Dataset: Simulation data	Framework : Distributed FL; Features : Ultrareliable communication, Low-latency design, Distributed learning	Privacy: Communication and vehicle privacy; Utility: Enhanced reliability (comparable accuracy with 79% reduction in data); Efficiency: Reduced latency and power consumption

(Continued)

	Blockchain-based PI	^{o}ML
--	---------------------	----------

Application: IoV block-streaming service: Data Type: User data, QoS data; Dataset: Simulation

Application: Toll pricing, Traffic prediction; [215] Data Type: Traffic and toll data; Dataset: New York State Thruway Authority dataset

Application: Fair energy trading; Data Type: [216] V2G energy transaction data; Dataset: Generated

Application: Security of SDN controller; Data [217] Type: Fog computing and 5G network data; Dataset: Simulation data

Application: Secure network topology sharing; [218] **Data Type**: Vehicular service offloading data; Dataset: Generated

Application: Secure data sharing; **Data Type**: [219]* Vehicle trace points, edge data (image; Dataset: Uber pickups in New York City, MNIST dataset Application: Knowledge sharing; Data Type:

[220] IoV environmental and knowledge-sharing image data; Dataset: MNIST and CIFAR10 dataset

Application: Secure SVM training; Data Type: [221] Vehicular social network data; Dataset: BCWD and ACAD dataset

Application: Secure vehicular communication; [222]* Data Type: Autonomous vehicle data; Dataset: Generated

Application: Message dissemination systems; [195]* **Data Type**: Vehicular communication data; Dataset: Generated

Framework: Blockchain-FL; Features: Data chunking for low-latency, Anonymization; Edge caching mechanism, PoW + PBFT

Framework: DwaRa framework; Features: Privacy budget optimization, Dynamic toll pricing, Privacy budget optimization

Framework: V2GEx framework; Features: Zero-knowledge funds, Hashchain micropayment, Smart contracts, PoW

Framework: Blockchain-SDN; Features: Fog computing, Blockchain-SDN integration, Distributed trust management, PoW and PoS

Framework: VEC trust management; Features: Service offloading, Vehicle migration, Trust computation, PBFT consensus

Framework: Hybrid blockchain and Asynchronous FL; Features: Auto model validation, DRL optimization, 2-stage verification convergence, Efficient data sharing

Framework: Hierarchical blockchain; Features: Multi-level architecture, Cross-domain learning, Smart contracts, PoK consensus

Framework: Consortium blockchain; Features: Vertical partitioning, Secure computation, Social

Framework: BFL framework; Features: Renewal reward approach, Block optimization, Consensus mechanism, PoW and PoS consensus

Framework: Blockchain-FL; Features: Secure dissemination. Blockchain verification. Distributed consensus, PoFL consensus

Privacy: User privacy, IoV device and edge node verification; Utility: Improved cache hit rate; Efficiency: Lower energy consumption and delay

Privacy: Data protection; Utility: Dynamic pricing, 0.0012 MSE; Efficiency: Lowered time (45.88ms) and communication cost (53bytes)

Privacy: Transaction anonymity, Identity privacy; **Utility**: Lower verification time (20ms); Efficiency: Low latency of 6s

Privacy: Network isolation, Access control; Utility: High throughput; Efficiency: Low latency

Privacy: Service offloading privacy; Utility: Service optimization, high throughput; Efficiency: Minimize delay and energy usage

Privacy: Resistance to tampering, Secure update; Utility: High mode accuracy; Efficiency: Fast

Privacy: Layer-wise privacy, Knowledge isolation; Utility: 10% better accuracy; Efficiency: Efficient knowledge transfer

Privacy: Feature privacy, Training data protection; Utility: Accurate SVM classifier; Efficiency: Low time cost, High communication

Privacy: Model security, Training privacy; Utility: Adaptive design, optimal block arrival rate; Efficiency: Minimized delay

Privacy: Message privacy; Utility: Network reliability, high accuracy; Efficiency: 65.2% faster, 8.2% more efficient dissemination

3. Homomorphic Encryption:

Application: Matching for EV Charging; Data [223] **Type**: EV charging demand and supplier data; Dataset: Generated

Application: Dynamic pricing in V2G Networks; [224] Data Type: network data; Dataset: Generated

Application: EV charging demand prediction; [225]* Data Type: Energy demand data; Dataset: Dundee EV Charging Sessions, CAN dataset

Application: Ride-sharing; Data Type: Ridesharing route and user query data; Dataset: [226] Generated

Application: Toll systems for congestion automation; Data Type: Toll transponder, vehicle location data; Dataset: From UHF RFID

Framework: Privacy-aware matching; Features: Privacy: Location anonymity, Identity Secure pairing protocol, Location protection, Efficient matching

Framework: PADP framework; Features: SASD aggregation, Threshold Paillier HE, Dynamic pricing aggregation

Framework: P3 framework; Features: CNN-BiLSTM model, CKKS cryptosystem, Blockchain integration

Framework: HE-based routing; Features: Paillier encryption, Private routing matching, Secure computation

Framework: TollsOnly framework; Features: Post-quantum HE, GDPR compliance, Blockchain integration

protection; Utility: Optimal matching;

Efficiency: Linear complexity, low waiting time

Privacy: price privacy, k-threshold security; Utility: Fair pricing, prevent impersonation attack; Efficiency: Minimized overhead

Privacy: Parameter security; Utility: High prediction accuracy; Efficiency: Low latency and computation

Privacy: Route privacy; Utility: Strong privacy and security guarantees; Efficiency: High computation cost

Privacy: Transponder privacy, user control; Utility: privacy risk assessment, monetize driving data; Efficiency: N/A

4. Secure Multi-party Computation:

Application: Vehicular data protection in IoV; Data Type: IoV data; Dataset: Generated

Application: Secure vehicular path tracking; [232] Data Type: Spatio-temporal maps; Dataset: Beijing taxis dataset

Application: CAV data validation; Data Type: Mobility data from CAVs and micro-mobility [233] devices; Dataset: New York City Connected Vehicles Pilot Study dataset

Application: Collaborative filtering in IVI; Data [234]* Type: VANET user interest and location data; Dataset: Simulation Data

Framework: AI-blockchain; Features: Multiparty computation, Auto-coding features, Decentralized security, -Intelligent contracts

Framework: Game theory-based framework; Features: Incentive mechanism, Nash equilibrium, SMPC integration

Framework: BELIEVE framework; Features: Blockchain-based MPC, Smart contract validation, IPFS integration

Framework: Cloud-based CF; Features: Homomorphic encryption, Interest-based sorting, Flexible updates

Privacy: transaction privacy; Utility: higher security utility, below 80% accuracy; Efficiency: Improved transaction verification and energy use

Privacy: Data anonymity, Path privacy; Utility: Prevent adversarial attack, incentivizes vehicle participation; Efficiency: Numerical instability

Privacy: Real-time privacy, Data integrity; **Utility**: secure validation and storage of mobility data; Efficiency: Resource optimization, Low delay (7 µs for 50 nodes)

Privacy: User privacy, data freshness; Utility: secure collaborative filtering; Efficiency: Low time and communication cost

(Continued)

5. Differential Privacy:

Application: EV charging; Data Type: EV charging data; Dataset: ACN-Data dataset

Application: EV querying; Data Type: EV [236] charging station query data; **Dataset**: OpenStreetMap dataset

Application: Parking systems, Charging station; [237] Data Type: Parking recommendation data; Dataset: Santander Spain parking dataset

Application: Secure communication, Traffic [238]* flow estimation; **Data Type**: VANET communication data; Dataset: i-VANET dataset

[194]* Application: Vehicular CPS; Data Type: VCPS image data; Dataset: MNIST dataset

Framework: Variable window DP: Features: Adaptive window sizing, Dynamic budget, Data utility preservation

Framework: AGeoI framework; Features: Approximate geo-indistinguishability, Dummy data generation, Bayesian updates

Framework: ECC-LDP framework; Features: Elliptic curve crypto, HMAC authentication, Laplace noise.

Framework: FL with LDP; Features: Hybrid FL Privacy: Data privacy; Utility: High reliability architecture, Local differential privacy, Distributed learning

Framework: FL with DP: Features: Distributed ML, Edge computing integration, Resilient aggregation

Privacy: User privacy; Utility: reduced interference errors and publishing errors; Efficiency: Optimal privacy budget allocation Privacy: Location privacy; Utility: High QoS maintained; Efficiency: "Privacy-for-free" for

Privacy: Location privacy, Query privacy; Utility: High recommendation accuracy; Efficiency: Low storage overheads, computation, and communication

against inference and gradient leakage attacks; Efficiency: Resource optimization, Dast training

Privacy: Adversarial resistance. Model privacy: Utility: Low accuracy drop for high privacy budget; Efficiency: Enhanced resilience

Notation: * Indicate integration of more than one PPML in the study (Hybrid PPML).

transmission overhead, and computation efficiency. Additionally, while many existing PPML techniques focus on embedding privacy into specific ML frameworks, there is no universal consensus on privacy guarantees, especially regarding threat models or trust assumptions. Achieving a standardized definition of privacy guarantees remains a significant challenge. We have divided PPML approaches into three categories as discussed in Section. III.B and analyzed the challenges associated with each specifically.

1) Architecture-based PPML Approaches: Architectural PPML frameworks, such as FL and BC-PPML systems, decentralize computation to avoid centralized aggregation. FL enables collaborative model training across distributed clients while retaining data locality, whereas blockchain ensures auditability and consensus via immutable ledgers. However, these systems face systemic vulnerabilities. iterative model update mechanism introduces communication inefficiencies, particularly in large-scale deployments with non-IID data distributions [239]. Non-IID data skews local client updates, degrading global model convergence and fairness. Moreover, FL gradients, though designed to protect raw data, remain susceptible to inference attacks such as membership inference and model inversion [240], [241]. Recent studies show these attacks require minimal assumptions, succeeding even in black-box settings where adversaries only access model APIs [242].

Poisoning attacks further threaten architectural PPML. Clean-label poisoning subtly alters training data without modifying labels, while dirty-label poisoning injects mislabeled samples. Model poisoning can achieve high attack success rates even with minimal poisoned data [243]. Byzantine attacks, such as uploading malicious gradients, exploit FL's aggregation protocols. Defensive mechanisms like Krum [244] and Trimmed Mean [245] partially mitigate these risks but struggle with scalability and computational costs. System heterogeneity, including variable client hardware and network conditions, further complicates uniform privacy integration.

To address these challenges, hybrid frameworks combining FL with blockchain could enhance trust and auditability. Lightweight consensus protocols (like Proof-of-Authority) may reduce blockchain latency, while gradient compression (sparsification, quantization) and adaptive client selection algorithms can mitigate FL's communication overhead [246]. Trusted execution environments (TEEs) like Intel SGX [247] could secure aggregation processes, and Byzantine-resilient techniques (such as gradient clipping) may neutralize poisoned updates [92]. Tokenized incentive systems, embedded via blockchain smart contracts, could incentivize honest participation. Future research must also refine privacy-utility trade-offs when integrating DP or SMPC into FL workflows. For example, DP noise injection during gradient aggregation reduces privacy leakage but degrades model accuracy, necessitating adaptive budget allocation strategies tailored to non-IID settings.

2) Data Processing-based PPML Approaches: Data processing techniques, including HE and SMPC, enable computation on encrypted or partitioned data. SMPC protocols like garbled circuits and secret sharing distribute computations across parties without revealing private inputs. However, these methods incur significant overheads. Garbled circuits encode Boolean logic operations via permuted truth tables, requiring multi-round peer-to-peer communication and quadratic scaling with model complexity. Pairwise masking-based secure aggregation, common in SMPC, further strains scalability for DNNs [18]. HE allows arithmetic operations on ciphertexts but struggles with non-linear functions (like ReLU) due to polynomial approximations [17]. Moreover, HE's reliance on lattice-based cryptography introduces latency from ciphertext expansion, especially in DNN inference [17].

A critical challenge lies in balancing encoding precision and computational efficiency. Most HE schemes (like CKKS) encode floating-point numbers into ciphertexts, but lower precision accelerates computation at the cost of accuracy. For instance, reducing mantissa bits in CKKS encoding speeds up homomorphic convolutions but introduces rounding errors that degrade performance [248]. Similarly, SMPC protocols require custom circuit designs for each task, limiting flexibility. Recent work on hybrid HE-SMPC frameworkswhere linear layers are computed under HE and non-linear activations under SMPC—offers a promising trade-off [18].

advancements must prioritize cryptographic optimizations and hardware acceleration. GPU-accelerated HE libraries tailored for CKKS/BFV schemes could expedite polynomial encrypted inference, while Chebyshev approximations may enable HE-compatible ReLU activations [17]. Reducing garbled circuit gates via automated circuit optimization tools (e.g., TinyGarble) and parallelizing operations could enhance scalability in SMPC integration [18]. Hardware-software co-design, such as integrating HE operations into AI accelerators (such as TPUs) or leveraging TEEs for secure SMPC coordination, may further reduce latency [18]. Standardized benchmarks for cryptographic ML workloads-evaluating latency, communication, and privacy guarantees—will guide protocol selection. Emerging neural processing units (NPUs), like Google's TPU or NVIDIA's NVDLA [249] could offload HE/SMPC computations, though their efficacy in non-neural tasks (e.g., cryptographic primitives) remains underexplored.

3) Data Publishing-based PPML Approaches: Data publishing techniques, such as DP, ensure statistical privacy by injecting calibrated noise into datasets or model outputs. The privacy budget (ϵ) governs the trade-off between privacy and utility: smaller ϵ values strengthen privacy but degrade model accuracy, while larger ϵ increases vulnerability to inference attacks [14], [27]. In distributed settings, local DP mechanisms protect individual contributions but complicate global privacy management. For example, FL workflows require careful composition of per-client ϵ budgets to prevent budget exhaustion during iterative training. Nonconvex optimization in modern DL models exacerbates these issues, as DP noise disrupts gradient descent trajectories, leading to suboptimal minima [12] [250].

Furthermore, recent studies highlight the tension between DP and fairness. Noise injection disproportionately impacts underrepresented groups in skewed datasets, amplifying biases. Adaptive DP mechanisms, such as per-feature noise scaling or gradient-specific clipping, could mitigate this by dynamically adjusting noise based on data sensitivity [251]. Integrating DP with synthetic data generation (like DP-GANs) or dimensionality reduction (like autoencoders) may preserve utility in high-dimensional spaces. Advanced composition frameworks, such as Rényi DP or zero-concentrated DP [252], [253], offer tighter privacy bounds for iterative workflows, enabling longer training without budget exhaustion.

TEEs such as INTEL SGX [247] and ARM TrustZone [254], have been proposed for privacy-preserving predictions. While TEEs isolate sensitive computations, implementation flaws (like side-channel leaks) limit their confidentiality guarantees. Future research could refine TEE architectures to resist physical and timing attacks, enabling secure DP noise generation or model aggregation. Compression techniques, such as neural network pruning and quantization, reduce computational costs for DP-based training. For instance, [255] demonstrated that pruning and Huffman coding can compress models by 35x–49x without accuracy loss, making DP

workflows more feasible for resource-constrained clients.

B. PPML in IoV Application Domains

The integration of PPML into the IoV ecosystem presents a complex interplay of technical, and infrastructural challenges, necessitating domain-specific solutions across the key application domains. As IoV systems increasingly rely on distributed data sources-from vehicular sensors RSUs to cloud-based analytics—the adoption of PPML techniques must contend with the unique constraints of real-time processing, heterogeneous data interoperability, and stringent safety requirements. While these techniques offer promising avenues to safeguard sensitive vehicular and user data, their deployment in latency-critical, safety-driven IoV environments introduces fundamental trade-offs between privacy guarantees and system performance. This section discusses these challenges across the three key IoV application domains identifying domain-specific barriers to PPML integration. For each domain, we first analyze challenges from computational overheads, protocol interoperability, and context-aware privacy-preservation, followed by research directions aimed at optimizing PPML architectures for the dynamic and large-scale IoV ecosystem.

1) Intelligent Transportation and Traffic Management: The adoption of PPML in intelligent transportation systems faces scalability-efficiency-privacy trilemmas. FL and SMPC incur significant communication overheads when aggregating heterogeneous data (like LiDAR point clouds, camera feeds, and IoT sensor streams) from millions of vehicles and RSUs [18], [20], [51], [170]. These protocols struggle to reconcile privacy preservation with the low latency demands of realtime applications like traffic flow prediction and adaptive signal control. HE, though theoretically robust, introduces prohibitive latency in large-scale traffic simulations due to polynomial-degree ciphertext operations, conflicting with subsecond response requirements for dynamic traffic management [17], [26], [64]. DP integration also exacerbate challenges in regions with sparse vehicular data, where uniform noise injection can amplify biases in traffic forecasting models, leading to inequitable and inaccurate decisions.

To address these challenges, hybrid PPML frameworks could leverage lightweight HE variants (like CKKS for approximate arithmetic) with hierarchical FL architectures [17], [99], [220], distributing computation across edge devices, RSUs, and cloud servers to minimize latency. Geospatial adaptive DP mechanisms could dynamically calibrate noise levels based on regional data density, preserving privacy in sparse zones while maintaining accuracy in dense traffic areas [256]. Blockchain-based FL systems can enhance trust in multi-jurisdictional traffic management by providing immutable audit trails for model updates via quantum-resistant lattice-based signatures (like the CRYSTALS-Dilithium [257]).

2) Autonomous Driving and Safety-critical Applications: Safety-critical autonomous systems require robust privacy guarantees without compromising real-time performance.

However, current PPML techniques fall short in meeting these demands. For instance, HE and SMPC-based collision avoidance systems may face a latency-security contradiction. Encrypted computations for multi-sensor fusion (LiDAR, radar, camera) introduce millisecond-level jeopardizing emergency braking or pedestrian detection. DP can degrade perception model precision—as noise in general can reduce detection or bounding-box accuracy [258], [259], increasing false negatives in cluttered environments. The opacity of PPML frameworks also conflicts with automotive safety standards (e.g., ISO 26262), which emphasize the importance of traceability and documentation throughout the safety lifecycle of automotive electronic/electrical systems [260]. For instance, encrypted inference pipelines obscure saliency maps, complicating forensic analysis in autonomous systems or accidents.

Hardware-accelerated PPML architectures (e.g. sparse homomorphic convolution kernels) [261], could mitigate latency bottlenecks in encrypted sensor fusion. Context-aware DP frameworks can employ advanced adaptive techniques [251], for noise budgets based on environmental risk— like reducing noise in certain areas and increasing noise in sensitive zones. To reconcile privacy with explainability, interpretable PPML models could integrate privacy-preserving attention mechanisms [262] or explainable AI [263], enabling auditable decisions without exposing raw sensor data. Additionally, federated reinforcement learning (FRL) with DP guarantees [264] could enable collaborative, privacy-preserving training of autonomous policies across vehicle fleets while adhering to safety-critical latency constraints.

3) Communication Infrastructure and Smart Services: Deploying PPML in IoV infrastructure and smart services faces interoperability, sustainability, and legacy compatibility issues. Conflicting PPML protocols (e.g., FL aggregation rules vs. blockchain consensus mechanisms) across service provides can impede secure data exchange in V2X networks, limiting scalability for smart CSs, grid-balancing algorithms, and maintenance systems. Retrofitting predictive legacy infrastructure—such as aging RSUs or traffic control systems—with modern PPML techniques like HE or SMPC is challenging due to limited computational resources. Furthermore, the bidirectional V2G ecosystem introduces unique privacy risks. Centralized training on EV charging patterns risks exposing user habits, while blockchain's transparency allows adversaries to infer participant behaviors from energy auction histories. Vehicular cloud platforms, crucial for applications like collaborative event sharing and anomaly detection, struggle to balance utility with privacymodels trained on cloud-stored data may inadvertently leak attributes of legitimate users [7]. Cognitive radio integration, though promising for dynamic spectrum allocation, risks exposing EV mobility trends through raw spectrum usage data [265], [266]. Edge computing and UAV-assisted blockchain networks, despite improving computational efficiency [30], face reliability issues such as untrusted edge servers may compromise vehicle trajectory data, while UAV mobility and

IRS reflection-angle dependencies degrade communication stability in high-density urban environments [30].

To address interoperability, standardized PPML interfaces could harmonize FL, HE, and blockchain protocols across V2X ecosystems. For legacy systems, modular PPML toolkits with hardware abstraction layers could enable incremental upgrades— such as deploying HE-enabled FPGAs for encrypted toll calculations or DP-enhanced edge gateways for privacy-aware traffic monitoring. In V2G networks, hybrid FL-DP frameworks could decentralize load forecasting by injecting calibrated noise [251] at local aggregators before global model training. Furthermore, stealth addresses and zk-SNARKs [267] could anonymize blockchain transaction. Cognitive radio systems can adopt FRL with DP to train spectrum allocation policies without exposing EV locations. For vehicular clouds, HE-based anomaly detection (like TFHE-encoded classifiers [268]) could secure encrypted data processing, and TEEs like Intel SGX [247] could isolate sensitive operations in intrusion detection models.

VI. CONCLUSION

The integration of ML into the IoV has significantly enhanced transportation efficiency and autonomous driving capabilities. However, it also introduces significant privacy risks due to the sensitivity of vehicular, environmental and user data. Architectural PPML techniques such as FL and BC-PPML enable privacy preserving decentralized collaboration, while computational PPML techniques such as HE, SMPC and DP protect data against adversarial attacks and unauthorized inference. This survey provided a comprehensive review of the recent advances in adopting PPML techniques for IoV applications. We systematically analyzed the privacy challenges inherent to ML-driven IoV systems, including sensitive data exposure, adversarial attacks, and communication vulnerabilities. We categorized IoV applications into three key domains and evaluated how PPML techniques effectively mitigate privacy risks while preserving utility. Despite various advancements, significant challenges remain, such as balancing privacy-utility trade-offs, managing computational overhead, and ensuring scalability across heterogeneous networks. To overcome these, we further discussed potential future directions such as hybrid PPML frameworks combining multiple techniques, and lightweight encryption for edge devices, among others.

REFERENCES

- O. Kaiwartya et al., "Internet of Vehicles: Motivation, Layered Architecture, Network Model, Challenges, and Future Aspects," *IEEE Access*, vol. 4, pp. 5356–5373, 2016.
- [2] F. Yang, S. Wang, J. Li, Z. Liu, and Q. Sun, "An overview of Internet of Vehicles," *China Commun.*, vol. 11, no. 10, pp. 1–15, Oct. 2014.
- [3] J. Contreras-Castillo, S. Zeadally, and J. A. Guerrero-Ibanez, "Internet of Vehicles: Architecture, Protocols, and Security," *IEEE Internet Things J.*, vol. 5, no. 5, pp. 3701–3709, Oct. 2018.
- [4] Z. Lu, G. Qu, and Z. Liu, "A Survey on Recent Advances in Vehicular Network Security, Trust, and Privacy," *IEEE Trans. Intell. Transport.* Syst., vol. 20, no. 2, pp. 760–776, Feb. 2019.
- [5] E. S. Ali et al., "Machine Learning Technologies for Secure Vehicular Communication in Internet of Vehicles: Recent Advances and

- Applications," Security and Communication Networks, vol. 2021, pp. 1–23, Mar. 2021.
- [6] A. Talpur and M. Gurusamy, "Machine learning for security in vehicular networks: A comprehensive survey," *IEEE Communications* Surveys & Tutorials, vol. 24, no. 1, pp. 346–379, 2021.
- [7] M. Rigaki and S. Garcia, "A Survey of Privacy Attacks in Machine Learning," ACM Comput. Surv., vol. 56, no. 4, pp. 1–34, Apr. 2024.
- [8] R. Xu, N. Baracaldo, and J. Joshi, "Privacy-Preserving Machine Learning: Methods, Challenges and Directions," Sep. 22, 2021, arXiv: arXiv:2108.04417.
- [9] F. Mireshghallah, M. Taram, P. Vepakomma, A. Singh, R. Raskar, and H. Esmaeilzadeh, "Privacy in Deep Learning: A Survey," Nov. 07, 2020, arXiv: arXiv:2004.12254.
- [10] A. Boulemtafes, A. Derhab, and Y. Challal, "A review of privacy-preserving techniques for deep learning," *Neurocomputing*, vol. 384, pp. 21–45, 2020.
- [11] E. Antwi-Boasiako, S. Zhou, Y. Liao, Q. Liu, Y. Wang, and K. Owusu-Agyemang, "Privacy preservation in Distributed Deep Learning: A survey on Distributed Deep Learning, privacy preservation techniques used and interesting research directions," *Journal of Info Security and Applications*, vol. 61, p. 102949, 2021.
- [12] Z. Chen and Y. Wang, "Privacy-Preserving Distributed Optimization and Learning," Feb. 29, 2024, arXiv: arXiv:2403.00157.
- [13] A.-T. Tran, T.-D. Luong, and V.-N. Huynh, "A comprehensive survey and taxonomy on privacy-preserving deep learning," *Neurocomputing*, p. 127345, 2024.
- [14] M. U. Hassan, M. H. Rehmani, and J. Chen, "Differential privacy techniques for cyber physical systems: A survey," *IEEE Comm. Surveys & Tutorials*, vol. 22, no. 1, pp. 746–789, 2019.
- [15] M. U. Hassan, M. H. Rehmani, and J. Chen, "Privacy preservation in blockchain based IoT systems: Integration issues, prospects, challenges, and future research directions," Future Generation Computer Systems, vol. 97, pp. 512–529, 2019.
- [16] A. El Ouadrhiri and A. Abdelhadi, "Differential privacy for deep and federated learning: A survey," *IEEE access*, vol. 10, pp. 22359– 22380, 2022.
- [17] B. Pulido-Gaytan et al., "Privacy-preserving neural networks with Homomorphic encryption: Challenges and opportunities," Peer-to-Peer Netw. Appl., vol. 14, no. 3, pp. 1666–1691, May 2021.
- [18] Q. Zhang, C. Xin, and H. Wu, "Privacy-Preserving Deep Learning Based on Multiparty Secure Computation: A Survey," *IEEE Internet Things J.*, vol. 8, no. 13, pp. 10412–10429, Jul. 2021.
- [19] X. Yin, Y. Zhu, and J. Hu, "A Comprehensive Survey of Privacy-preserving Federated Learning: A Taxonomy, Review, and Future Directions," ACM Comput. Surv., vol. 54, no. 6, pp. 1–36, Jul. 2022.
- [20] E. T. M. Beltrán et al., "Decentralized federated learning: Fundamentals, state of the art, frameworks, trends, and challenges," IEEE Communications Surveys & Tutorials, 2023.
- [21] Q. Han, S. Lu, W. Wang, H. Qu, J. Li, and Y. Gao, "Privacy preserving and secure robust federated learning: A survey," *Concurrency and Computation*, vol. 36, no. 13, p. e8084, Jun. 2024.
- [22] D. Hahn, A. Munir, and V. Behzadan, "Security and privacy issues in intelligent transportation systems: Classification and challenges," *IEEE Intelligent Transportation Systems Magazine*, vol. 13, no. 1, pp. 181–196, 2019.
- [23] L. Sleem, H. N. Noura, and R. Couturier, "Towards a secure ITS: Overview, challenges and solutions," *Journal of Information Security and Applications*, vol. 55, p. 102637, 2020.
- [24] A. R. Sani, M. U. Hassan, L. Gao, and J. Chen, "Privacy Preserving Machine Learning for Electric Vehicles: A Survey," Dec. 20, 2024, arXiv: arXiv:2205.08462.
- [25] D. P. Moya Osorio et al., "Towards 6G-Enabled Internet of Vehicles: Security and Privacy," *IEEE Open J. Commun. Soc.*, vol. 3, pp. 82–105, 2022.
- [26] X. Sun, F. R. Yu, P. Zhang, W. Xie, and X. Peng, "A Survey on Secure Computation Based on Homomorphic Encryption in Vehicular Ad Hoc Networks," Sensors, vol. 20, no. 15, p. 4253, Jul. 2020.
- [27] P. Zhao, G. Zhang, S. Wan, G. Liu, and T. Umer, "A survey of local differential privacy for securing internet of vehicles," *J Supercomput*, vol. 76, no. 11, pp. 8391–8412, Nov. 2020.
- [28] K. Kaltakis, P. Polyzi, G. Drosatos, and K. Rantos, "Privacy-Preserving Solutions in Blockchain-Enabled Internet of Vehicles," Applied Sciences, vol. 11, no. 21, p. 9792, Oct. 2021.

- [29] M. Billah, Sk. T. Mehedi, A. Anwar, Z. Rahman, and R. Islam, "A Systematic Literature Review on Blockchain Enabled Federated Learning Framework for Internet of Vehicles," 2022, arXiv.
- [30] X. Wang, H. Zhu, Z. Ning, L. Guo, and Y. Zhang, "Blockchain Intelligence for Internet of Vehicles: Challenges and Solutions," *IEEE Commun. Surv. Tutorials*, vol. 25, no. 4, pp. 2325–2355, 2023.
- [31] V. P. Chellapandi, L. Yuan, C. G. Brinton, S. H. Żak, and Z. Wang, "Federated Learning for Connected and Automated Vehicles: A Survey of Existing Approaches and Challenges," *IEEE Trans. Intell.* Veh., vol. 9, no. 1, pp. 119–137, Jan. 2024.
- [32] S. Zhang *et al.*, "Federated Learning in Intelligent Transportation Systems: Recent Applications and Open Problems," *IEEE Trans. Intell. Transport. Syst.*, vol. 25, no. 5, pp. 3259–3285, May 2024.
- [33] R. Zhang, J. Mao, H. Wang, B. Li, X. Cheng, and L. Yang, "A Survey on Federated Learning in Intelligent Transportation Systems," *IEEE Trans. Intell. Veh.*, pp. 1–17, 2024.
- [34] S. Tuohy, M. Glavin, C. Hughes, E. Jones, M. Trivedi, and L. Kilmartin, "Intra-Vehicle Networks: A Review," *IEEE Trans. Intell. Transport. Syst.*, vol. 16, no. 2, pp. 534–545, Apr. 2015.
- [35] J. Wang, J. Liu, and N. Kato, "Networking and Communications in Autonomous Driving: A Survey," *IEEE Commun. Surv. Tutorials*, vol. 21, no. 2, pp. 1243–1274, 2019.
- [36] H. Taslimasa, S. Dadkhah, E. C. P. Neto, P. Xiong, S. Ray, and A. A. Ghorbani, "Security issues in Internet of Vehicles (IoV): A comprehensive survey," *Internet of Things*, vol. 22, p. 100809, Jul. 2023
- [37] K. N. Qureshi, S. Din, G. Jeon, and F. Piccialli, "Internet of Vehicles: Key Technologies, Network Model, Solutions and Challenges With Future Aspects," *IEEE Trans. Intell. Transport. Syst.*, vol. 22, no. 3, pp. 1777–1786, Mar. 2021.
- [38] R. Ravindran, M. J. Santora, and M. M. Jamali, "Multi-Object Detection and Tracking, Based on DNN, for Autonomous Vehicles: A Review," *IEEE Sensors J.*, vol. 21, no. 5, pp. 5668–5677, Mar. 2021.
- [39] A. K. Haghighat, V. Ravichandra-Mouli, P. Chakraborty, Y. Esfandiari, S. Arabi, and A. Sharma, "Applications of Deep Learning in Intelligent Transportation Systems," *J. Big Data Anal. Transp.*, vol. 2, no. 2, pp. 115–145, Aug. 2020.
- [40] S. Porru, F. E. Misso, F. E. Pani, and C. Repetto, "Smart mobility and public transport: Opportunities and challenges in rural and urban areas," *Journal of Traffic and Transportation Engineering (English Edition)*, vol. 7, no. 1, pp. 88–97, Feb. 2020.
- [41] E. Yurtsever, J. Lambert, A. Carballo, and K. Takeda, "A Survey of Autonomous Driving: Common Practices and Emerging Technologies," IEEE Access, vol. 8, pp. 58443–58469, 2020.
- [42] S. Atakishiyev, M. Salameh, H. Yao, and R. Goebel, "Explainable Artificial Intelligence for Autonomous Driving: A Comprehensive Overview and Field Guide for Future Research Directions," *IEEE Access*, vol. 12, pp. 101603–101625, 2024.
- [43] H. Zhou, W. Xu, J. Chen, and W. Wang, "Evolutionary V2X Technologies Toward the Internet of Vehicles: Challenges and Opportunities," *Proc. IEEE*, vol. 108, no. 2, pp. 308–323, Feb. 2020.
- [44] A. Hemmati, M. Zarei, and A. Souri, "UAV-based Internet of Vehicles: A systematic literature review," *Intelligent Systems with Applications*, vol. 18, p. 200226, May 2023.
- [45] J. Guo, M. Bilal, Y. Qiu, C. Qian, X. Xu, and K.-K. Raymond Choo, "Survey on digital twins for Internet of Vehicles: Fundamentals, challenges, and opportunities," *Digital Communications and Networks*, vol. 10, no. 2, pp. 237–247, Apr. 2024.
- [46] D. B. Rawat and C. Bajracharya, Vehicular Cyber Physical Systems. Cham: Springer International Publishing, 2017.
- [47] N. I. Nimalsiri, C. P. Mediwaththe, E. L. Ratnam, M. Shaw, D. B. Smith, and S. K. Halgamuge, "A Survey of Algorithms for Distributed Charging Control of Electric Vehicles in Smart Grid," *IEEE Trans. Intell. Transport. Syst.*, vol. 21, no. 11, pp. 4497–4515, Nov. 2020.
- [48] T. Lin, H. Rivano, and F. Le Mouel, "A Survey of Smart Parking Solutions," *IEEE Trans. Intell. Transport. Syst.*, vol. 18, no. 12, pp. 3229–3253, Dec. 2017.
- [49] R. Krstačić, A. Žužić, and T. Orehovački, "Safety Aspects of In-Vehicle Infotainment Systems: A Systematic Literature Review from 2012 to 2023," *Electronics*, vol. 13, no. 13, p. 2563, Jun. 2024.
- [50] W. Nai et al., "A Comprehensive Review of Driving Style Evaluation Approaches and Product Designs Applied to Vehicle Usage-Based Insurance," Sustainability, vol. 14, no. 13, p. 7705, Jun. 2022.

- [51] Z. Mahmood, "Connected Vehicles in the IoV: Concepts, Technologies and Architectures," in *Connected Vehicles in the Internet of Things*, Z. Mahmood, Ed., Cham: Springer International Publishing, 2020, pp. 3–18.
- [52] S. Arumugam and R. Bhargavi, "A survey on driving behavior analysis in usage based insurance using big data," *J Big Data*, vol. 6, no. 1, p. 86, Dec. 2019.
- [53] D. Oladimeji, K. Gupta, N. A. Kose, K. Gundogan, L. Ge, and F. Liang, "Smart Transportation: An Overview of Technologies and Applications," Sensors, vol. 23, no. 8, p. 3880, Apr. 2023.
- [54] A. S. Mohammed, A. Amamou, F. K. Ayevide, S. Kelouwani, K. Agbossou, and N. Zioui, "The Perception System of Intelligent Ground Vehicles in All Weather Conditions: A Systematic Literature Review," Sensors, vol. 20, no. 22, p. 6532, Nov. 2020.
- [55] Y. Zhang, A. Carballo, H. Yang, and K. Takeda, "Perception and sensing for autonomous vehicles under adverse weather conditions: A survey," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 196, pp. 146–177, Feb. 2023.
- [56] G. Abdelkader, K. Elgazzar, and A. Khamis, "Connected Vehicles: Technology Review, State of the Art, Challenges and Opportunities," Sensors, vol. 21, no. 22, p. 7712, Nov. 2021.
- [57] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Artificial intelligence and statistics*, PMLR, 2017, pp. 1273–1282.
- [58] L. Collins, H. Hassani, A. Mokhtari, and S. Shakkottai, "Fedavg with fine tuning: Local updates lead to representation learning," *Advances in Neural Info. Processing Systems*, vol. 35, pp. 10572–10586, 2022.
- [59] G. Hinton, O. Vinyals, and J. Dean, "Distilling the Knowledge in a Neural Network," 2015, arXiv.
- [60] A. Mora, I. Tenison, P. Bellavista, and I. Rish, "Knowledge Distillation for Federated Learning: a Practical Guide," Nov. 09, 2022, arXiv: arXiv:2211.04742.
- [61] Y. Liu, F. R. Yu, X. Li, H. Ji, and V. C. M. Leung, "Blockchain and Machine Learning for Communications and Networking Systems," *IEEE Commun. Surv. Tutorials*, vol. 22, no. 2, pp. 1392–1431, 2020.
- [62] N. Waheed, X. He, M. Ikram, M. Usman, S. S. Hashmi, and M. Usman, "Security and Privacy in IoT Using Machine Learning and Blockchain: Threats and Countermeasures," ACM Comput. Surv., vol. 53, no. 6, pp. 1–37, Nov. 2021.
- [63] R. L. Rivest, L. Adleman, and M. L. Dertouzos, "On data banks and privacy homomorphisms," *Foundations of secure computation*, vol. 4, no. 11, pp. 169–180, 1978.
- [64] A. Acar, H. Aksu, A. S. Uluagac, and M. Conti, "A Survey on Homomorphic Encryption Schemes: Theory and Implementation," ACM Comput. Surv., vol. 51, no. 4, pp. 1–35, Jul. 2019.
- [65] E. Hesamifard, H. Takabi, and M. Ghasemi, "CryptoDL: Deep Neural Networks over Encrypted Data," Nov. 14, 2017, arXiv: arXiv:1711.05189.
- [66] S. Lee, G. Lee, J. W. Kim, J. Shin, and M.-K. Lee, "HETAL: efficient privacy-preserving transfer learning with homomorphic encryption," in *International Conference on Machine Learning*, PMLR, 2023, pp. 19010–19035.
- [67] Y. Lindell, "Secure multiparty computation for privacy preserving data mining," in *Encyclopedia of Data Warehousing and Mining*, IGI global, 2005, pp. 1005–1009.
- [68] C. Zhao et al., "Secure Multi-Party Computation: Theory, practice and applications," *Information Sciences*, vol. 476, pp. 357–372, Feb. 2019.
- [69] R. Kumar, P. Kumar, R. Tripathi, G. P. Gupta, N. Kumar, and M. M. Hassan, "A Privacy-Preserving-Based Secure Framework Using Blockchain-Enabled Deep-Learning in Cooperative Intelligent Transport System," *IEEE Trans. Intell. Transport. Syst.*, vol. 23, no. 9, pp. 16492–16503, Sep. 2022.
- [70] H. Wang, R. Zhang, X. Cheng, and L. Yang, "Hierarchical Traffic Flow Prediction Based on Spatial-Temporal Graph Convolutional Network," *IEEE Trans. Intell. Transport. Syst.*, vol. 23, no. 9, pp. 16137–16147, Sep. 2022.
- [71] M. Xia, D. Jin, and J. Chen, "Short-Term Traffic Flow Prediction Based on Graph Convolutional Networks and Federated Learning," *IEEE Trans. Intell. Transport. Syst.*, vol. 24, no. 1, pp. 1191–1203, Jan. 2023.
- [72] C. Zhang, S. Zhang, J. J. Q. Yu, and S. Yu, "FASTGNN: A Topological Information Protected Federated Learning Approach for

- Traffic Speed Forecasting," *IEEE Trans. Ind. Inf.*, vol. 17, no. 12, pp. 8464–8474, Dec. 2021.
- [73] X. Yuan et al., "FedSTN: Graph Representation Driven Federated Learning for Edge Computing Enabled Urban Traffic Flow Prediction," *IEEE Trans. Intell. Transport. Syst.*, vol. 24, no. 8, pp. 8738–8748, Aug. 2023.
- [74] Y. Liu, J. J. Q. Yu, J. Kang, D. Niyato, and S. Zhang, "Privacy-Preserving Traffic Flow Prediction: A Federated Learning Approach," IEEE Internet Things J., vol. 7, no. 8, pp. 7751–7763, Aug. 2020.
- [75] C. Zhang, S. Zhang, S. Yu, and J. J. Q. Yu, "Graph-Based Traffic Forecasting via Communication-Efficient Federated Learning," in 2022 IEEE Wireless Communications and Networking Conference (WCNC), Austin, TX, USA: IEEE, Apr. 2022, pp. 2041–2046.
- [76] X. Yuan, J. Chen, N. Zhang, C. Zhu, Q. Ye, and X. S. Shen, "FedTSE: Low-Cost Federated Learning for Privacy-Preserved Traffic State Estimation in IoV," in *IEEE INFOCOM 2022 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, New York, NY, USA: IEEE, May 2022, pp. 1–6.
- [77] W. Wang, G. Yang, L. Bao, K. Ma, and H. Zhou, "A Privacy-Preserving Crowd Flow Prediction Framework Based on Federated Learning during Epidemics," *Security and Communication Networks*, vol. 2022, pp. 1–20, Oct. 2022.
- [78] F. Z. Errounda and Y. Liu, "A Mobility Forecasting Framework with Vertical Federated Learning," in 2022 IEEE 46th Annual Computers, Software, and Applications Conference (COMPSAC), Los Alamitos, CA, USA: IEEE, Jun. 2022, pp. 301–310.
- [79] P. W. Shaikh, M. El-Abd, M. Khanafer, and K. Gao, "A Review on Swarm Intelligence and Evolutionary Algorithms for Solving the Traffic Signal Control Problem," *IEEE Trans. Intell. Transport. Syst.*, vol. 23, no. 1, pp. 48–63, Jan. 2022.
- [80] X. Li and J.-Q. Sun, "Multi-objective optimal predictive control of signals in urban traffic network," *Journal of Intelligent Transportation Systems*, vol. 23, no. 4, pp. 370–388, Jul. 2019.
- [81] H. Jia, Y. Lin, Q. Luo, Y. Li, and H. Miao, "Multi-objective optimization of urban road intersection signal timing based on particle swarm optimization algorithm," *Advances in Mechanical Engineering*, vol. 11, no. 4, p. 1687814019842498, Apr. 2019.
- [82] T. Chu, J. Wang, L. Codeca, and Z. Li, "Multi-Agent Deep Reinforcement Learning for Large-Scale Traffic Signal Control," *IEEE Trans. Intell. Transport. Syst.*, vol. 21, no. 3, pp. 1086–1095, Mar. 2020.
- [83] T. Wang, T. Liang, J. Li, W. Zhang, Y. Zhang, and Y. Lin, "Adaptive Traffic Signal Control Using Distributed MARL and Federated Learning," in 2020 IEEE 20th International Conference on Communication Technology (ICCT), Nanning, China: IEEE, Oct. 2020, pp. 1242–1248.
- [84] Y. Ye, W. Zhao, T. Wei, S. Hu, and M. Chen, "FedLight: Federated Reinforcement Learning for Autonomous Multi-Intersection Traffic Signal Control," in 2021 58th ACM/IEEE Design Automation Conference (DAC), San Francisco, CA, USA: IEEE, Dec. 2021, pp. 847–852.
- [85] C. Li, Y. Zhang, and Y. Luo, "A Federated Learning-Based Edge Caching Approach for Mobile Edge Computing-Enabled Intelligent Connected Vehicles," *IEEE Trans. Intell. Transport. Syst.*, vol. 24, no. 3, pp. 3360–3369, Mar. 2023.
- [86] C. Xu and Y. Mao, "An Improved Traffic Congestion Monitoring System Based on Federated Learning," *Information*, vol. 11, no. 7, p. 365, Jul. 2020.
- [87] X. Kong, H. Gao, G. Shen, G. Duan, and S. K. Das, "FedVCP: A Federated-Learning-Based Cooperative Positioning Scheme for Social Internet of Vehicles," *IEEE Trans. Comput. Soc. Syst.*, vol. 9, no. 1, pp. 197–206, Feb. 2022.
- [88] Y. Zheng, Q. Li, Y. Chen, X. Xie, and W.-Y. Ma, "Understanding mobility based on GPS data," in *Proceedings of the 10th international* conference on Ubiquitous computing, Seoul Korea: ACM, Sep. 2008, pp. 312–321.
- [89] D. Vasisht, S. Kumar, and D. Katabi, "Decimeter-Level Localization with a Single WiFi Access Point," in 13th USENIX Symposium on Networked Systems Design and Implementation (NSDI 16), Santa Clara, CA: USENIX Association, Mar. 2016, pp. 165–178.
- [90] R. Gao et al., "Glow in the Dark: Smartphone Inertial Odometry for Vehicle Tracking in GPS Blocked Environments," *IEEE Internet Things J.*, vol. 8, no. 16, pp. 12955–12967, Aug. 2021.

- [91] Y. Zhu, Y. Liu, J. J. Q. Yu, and X. Yuan, "Semi-Supervised Federated Learning for Travel Mode Identification From GPS Trajectories," *IEEE Trans. Intell. Transport. Syst.*, vol. 23, no. 3, pp. 2380–2391, Mar. 2022.
- [92] Z. Li et al., "Byzantine Resistant Secure Blockchained Federated Learning at the Edge," *IEEE Network*, vol. 35, no. 4, pp. 295–301, Jul. 2021.
- [93] V. Hassija, V. Gupta, S. Garg, and V. Chamola, "Traffic Jam Probability Estimation Based on Blockchain and Deep Neural Networks," *IEEE Trans. Intell. Transport. Syst.*, vol. 22, no. 7, pp. 3919–3928, Jul. 2021.
- [94] Y. Qi, M. S. Hossain, J. Nie, and X. Li, "Privacy-preserving blockchain-based federated learning for traffic flow prediction," *Future Generation Comp. Systems*, vol. 117, pp. 328–337, Apr. 2021.
- [95] C. Meese, H. Chen, S. A. Asif, W. Li, C.-C. Shen, and M. Nejad, "BFRT: Blockchained Federated Learning for Real-time Traffic Flow Prediction," in 2022 22nd IEEE International Symposium on Cluster, Cloud and Internet Computing (CCGrid), Taormina, Italy: IEEE, May 2022, pp. 317–326.
- [96] H. Guo, C. Meese, W. Li, C.-C. Shen, and M. Nejad, "B² SFL: A Bi-Level Blockchained Architecture for Secure Federated Learning-Based Traffic Prediction," *IEEE Trans. Serv. Comput.*, vol. 16, no. 6, pp. 4360–4374, Nov. 2023.
- [97] S. M. Halim, L. Khan, and B. Thuraisingham, "Next Location Prediction Using Federated Learning on a Blockchain," in 2020 IEEE Second International Conference on Cognitive Machine Intelligence (CogMI), Atlanta, GA, USA: IEEE, Oct. 2020, pp. 244–250.
- [98] L. Lyu et al., "Towards Fair and Privacy-Preserving Federated Deep Models," *IEEE Trans. Parallel Distrib. Syst.*, vol. 31, no. 11, pp. 2524–2541, Nov. 2020.
- [99] X. Wang, S. Garg, H. Lin, G. Kaddoum, J. Hu, and M. M. Hassan, "Heterogeneous Blockchain and AI-Driven Hierarchical Trust Evaluation for 5G-Enabled Intelligent Transportation Systems," *IEEE Trans. Intell. Transport. Syst.*, pp. 1–10, 2021.
- [100] J. Kang et al., "Optimizing Task Assignment for Reliable Blockchain-Empowered Federated Edge Learning," *IEEE Trans. Veh. Technol.*, vol. 70, no. 2, pp. 1910–1923, Feb. 2021.
- [101] L. Feng, Z. Yang, S. Guo, X. Qiu, W. Li, and P. Yu, "Two-Layered Blockchain Architecture for Federated Learning Over the Mobile Edge Network," *IEEE Network*, vol. 36, no. 1, pp. 45–51, Jan. 2022.
- [102] J. Deng and G. Shen, "Federated Learning-based Privacy-Preserving Traffic Flow Prediction Scheme for VANETs," in 2022 4th International Conference on Communications, Information System and Computer Engineering (CISCE), Shenzhen, China: IEEE, May 2022, pp. 374–378.
- [103] Q. Kong et al., "Privacy-Preserving Aggregation for Federated Learning-Based Navigation in Vehicular Fog," *IEEE Trans. Ind. Inf.*, vol. 17, no. 12, pp. 8453–8463, Dec. 2021.
- [104] J. Chen, K. Li, and P. S. Yu, "Privacy-Preserving Deep Learning Model for Decentralized VANETs Using Fully Homomorphic Encryption and Blockchain," *IEEE Trans. Intell. Transport. Syst.*, vol. 23, no. 8, pp. 11633–11642, Aug. 2022.
- [105] G. Sun, S. Sun, J. Sun, H. Yu, X. Du, and M. Guizani, "Security and privacy preservation in fog-based crowd sensing on the internet of vehicles," *Journal of Network and Computer Applications*, vol. 134, pp. 89–99, May 2019.
- [106] Q. Kong, R. Lu, M. Ma, and H. Bao, "A privacy-preserving sensory data sharing scheme in Internet of Vehicles," *Future Generation Computer Systems*, vol. 92, pp. 644–655, Mar. 2019. future.2017.12.003.
- [107] J. Zhang, F. Yang, Z. Ma, Z. Wang, X. Liu, and J. Ma, "A Decentralized Location Privacy-Preserving Spatial Crowdsourcing for Internet of Vehicles," *IEEE Trans. Intell. Transport. Syst.*, vol. 22, no. 4, pp. 2299–2313, Apr. 2021.
- [108] J. Zhou, S. Chen, K.-K. R. Choo, Z. Cao, and X. Dong, "EPNS: Efficient Privacy Preserving Intelligent Traffic Navigation from Multiparty Delegated Computation in Cloud-Assisted VANETs," IEEE Trans. on Mobile Comput., pp. 1–1, 2021.
- [109] C. Tan and K. Yang, "Privacy-preserving adaptive traffic signal control in a connected vehicle environment," *Transportation Research Part C: Emerging Technologies*, vol. 158, p. 104453, Jan. 2024. trc.2023.104453.
- [110] M. Usman, M. A. Jan, and A. Jolfaei, "SPEED: A Deep Learning Assisted Privacy-Preserved Framework for Intelligent Transportation

- Systems," *IEEE Trans. Intell. Transport. Syst.*, vol. 22, no. 7, pp. 4376–4384, Jul. 2021.
- [111] M. Tsao, K. Yang, K. Gopalakrishnan, and M. Pavone, "Private Location Sharing for Decentralized Routing Services," in 2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC), Macau, China: IEEE, Oct. 2022, pp. 2479–2486.
- [112] M. Akallouch, O. Akallouch, K. Fardousse, A. Bouhoute, and I. Berrada, "Prediction and Privacy Scheme for Traffic Flow Estimation on the Highway Road Network," *Information*, vol. 13, no. 8, p. 381, Aug. 2022.
- [113] M. Gelderie, M. Luff, and L. Brodschelm, "Differential Privacy for Distributed Traffic Monitoring in Smart Cities:," in *Proceedings of the* 10th International Conference on Information Systems Security and Privacy, Rome, Italy: SCITEPRESS - Science and Technology Publications, 2024, pp. 758–765.
- [114] Y. Li, P. Zhang, and Y. Wang, "The Location Privacy Protection of Electric Vehicles with Differential Privacy in V2G Networks," *Energies*, vol. 11, no. 10, p. 2625, Oct. 2018.
- [115] S. Ghane, A. Jolfaei, L. Kulik, K. Ramamohanarao, and D. Puthal, "Preserving Privacy in the Internet of Connected Vehicles," *IEEE Trans. Intell. Transport. Syst.*, vol. 22, no. 8, pp. 5018–5027, Aug. 2021.
- [116] H. Mu, L. Yuan, and J. Li, "Human Sensing via Passive Spectrum Monitoring," *IEEE Open J. Instrum. Meas.*, vol. 2, pp. 1–13, 2023.
- [117] A. Ferrari, D. Micucci, M. Mobilio, and P. Napoletano, "Deep learning and model personalization in sensor-based human activity recognition," *J Reliable Intelligent Environment*, vol. 9, no. 1, pp. 27– 39, Mar. 2023.
- [118] R. Du, K. Han, R. Gupta, S. Chen, S. Labi, and Z. Wang, "Driver Monitoring-Based Lane-Change Prediction: A Personalized Federated Learning Framework," in 2023 IEEE Intelligent Vehicles Symposium (IV), Anchorage, AK, USA: IEEE, Jun. 2023, pp. 1–7.
- [119] Y. Lu et al., "A Shared Control Design for Steering Assistance System Considering Driver Behaviors," *IEEE Trans. Intell. Veh.*, vol. 8, no. 1, pp. 900–911, Jan. 2023.
- [120] Y. Ma et al., "M² DAR: Multi-View Multi-Scale Driver Action Recognition with Vision Transformer," in 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Vancouver, BC, Canada: IEEE, Jun. 2023, pp. 5287–5294.
- [121] K. Doshi and Y. Yilmaz, "Federated Learning-based Driver Activity Recognition for Edge Devices," in 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), New Orleans, LA, USA: IEEE, Jun. 2022, pp. 3337–3345.
- [122] C. He, M. Annavaram, and S. Avestimehr, "Group Knowledge Transfer: Federated Learning of Large CNNs at the Edge," in Advances in Neural Information Processing Systems, H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, Eds., Curran Associates, Inc., 2020, pp. 14068–14080.
- [123] L. Yuan, L. Su, and Z. Wang, "Federated Transfer–Ordered–Personalized Learning for Driver Monitoring Application," *IEEE Internet Things J.*, vol. 10, no. 20, pp. 18292–18301, Oct. 2023.
- [124] L. Zhang, H. Saito, L. Yang, and J. Wu, "Privacy-Preserving Federated Transfer Learning for Driver Drowsiness Detection," *IEEE Access*, vol. 10, pp. 80565–80574, 2022.
- [125] Q. Liu, Q. Guo, W. Wang, Y. Zhang, and Q. Kang, "An Automatic Detection Algorithm of Metro Passenger Boarding and Alighting Based on Deep Learning and Optical Flow," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–13, 2021.
- [126] L. Yuan, Y. Ma, L. Su, and Z. Wang, "Peer-to-Peer Federated Continual Learning for Naturalistic Driving Action Recognition," in 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Vancouver, BC, Canada: IEEE, Jun. 2023, pp. 5250–5259.
- [127] U. M. Gidado, H. Chiroma, N. Aljojo, S. Abubakar, S. I. Popoola, and M. A. Al-Garadi, "A Survey on Deep Learning for Steering Angle Prediction in Autonomous Vehicles," *IEEE Access*, vol. 8, pp. 163797–163817, 2020.
- [128] M. Bojarski et al., "End to End Learning for Self-Driving Cars," 2016, arXiv.
- [129] H. Zhang, J. Bosch, and H. H. Olsson, "End-to-End Federated Learning for Autonomous Driving Vehicles," in 2021 International Joint Conference on Neural Networks (IJCNN), Shenzhen, China: IEEE, Jul. 2021, pp. 1–8.

- [130] A. M P, G. R, and M. Panda, "Steering Angle Prediction for Autonomous Driving using Federated Learning: The Impact of Vehicle-To-Everything Communication," in 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT), Kharagpur, India: IEEE, Jul. 2021, pp. 1–7.
- [131] B. Paden, M. Cap, S. Z. Yong, D. Yershov, and E. Frazzoli, "A Survey of Motion Planning and Control Techniques for Self-Driving Urban Vehicles," *IEEE Trans. Intelligent Veh.*, vol. 1, no. 1, pp. 33– 55, Mar. 2016.
- [132] F. Tian, Z. Li, F.-Y. Wang, and L. Li, "Parallel Learning-Based Steering Control for Autonomous Driving," *IEEE Trans. Intell. Veh.*, vol. 8, no. 1, pp. 379–389, Jan. 2023.
- [133] S. Xiao, X. Ge, Q.-L. Han, and Y. Zhang, "Resource-Efficient Platooning Control of Connected Automated Vehicles Over VANETs," *IEEE Transaction on Intelligent Vehicles.*, vol. 7, no. 3, pp. 579–589, Sep. 2022.
- [134] T. Zeng, O. Semiari, M. Chen, W. Saad, and M. Bennis, "Federated Learning on the Road Autonomous Controller Design for Connected and Autonomous Vehicles," *IEEE Trans. Wireless Commun.*, vol. 21, no. 12, pp. 10407–10423, Dec. 2022.
- [135] T. Wu, M. Jiang, Y. Han, Z. Yuan, X. Li, and L. Zhang, "A Traffic-Aware Federated Imitation Learning Framework for Motion Control at Unsignalized Intersections with Internet of Vehicles," *Electronics*, vol. 10, no. 24, p. 3050, Dec. 2021.
- [136] S. Liu, Y. Fu, P. Zhao, F. Li, and C. Li, "Autonomous Braking Algorithm for Rear-End Collision via Communication-Efficient Federated Learning," in 2021 IEEE Global Comm. Conference (GLOBECOM), Madrid, Spain: IEEE, Dec. 2021, pp. 01–06.
- [137] S. Kuutti, R. Bowden, Y. Jin, P. Barber, and S. Fallah, "A Survey of Deep Learning Applications to Autonomous Vehicle Control," *IEEE Trans. Intell. Transport. Syst.*, vol. 22, no. 2, pp. 712–733, Feb. 2021.
- [138] S. Grigorescu, B. Trasnea, T. Cocias, and G. Macesanu, "A survey of deep learning techniques for autonomous driving," *Journal of Field Robotics*, vol. 37, no. 3, pp. 362–386, Apr. 2020.
- [139] X. Zhou, R. Ke, Z. Cui, Q. Liu, and W. Qian, "STFL:Spatio-temporal Federated Learning for Vehicle Trajectory Prediction," in 2022 IEEE 2nd International Conference on Digital Twins and Parallel Intelligence (DTPI), Boston, MA, USA: IEEE, Oct. 2022, pp. 1–6.
- [140] N. Majcherczyk, N. Srishankar, and C. Pinciroli, "Flow-FL: Data-Driven Federated Learning for Spatio-Temporal Predictions in Multi-Robot Systems," in 2021 IEEE Int. Conference on Robotics and Automation (ICRA), Xi'an, China: IEEE, May 2021, pp. 8836–8842.
- [141] C. Koetsier, J. Fiosina, J. N. Gremmel, J. P. Müller, D. M. Woisetschläger, and M. Sester, "Detection of anomalous vehicle trajectories using federated learning," ISPRS Open Journal of Photogrammetry and Remote Sensing, vol. 4, p. 100013, Apr. 2022.
- [142] M. Han, K. Xu, S. Ma, A. Li, and H. Jiang, "Federated learning-based trajectory prediction model with privacy preserving for intelligent vehicle," *Int J of Intelligent Sys*, vol. 37, no. 12, pp. 10861–10879, Dec. 2022.
- [143] G. Rjoub, J. Bentahar, and O. A. Wahab, "Explainable AI-based Federated Deep Reinforcement Learning for Trusted Autonomous Driving," in 2022 Int. Wireless Communications & Mobile Computing (IWCMC), Dubrovnik, Croatia: IEEE, May 2022, pp. 318–323.
- [144] C. Wang, X. Chen, J. Wang, and H. Wang, "ATPFL: Automatic Trajectory Prediction Model Design under Federated Learning Framework," in 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA: IEEE, Jun. 2022, pp. 6553–6562.
- [145] M. Nakanoya, J. Im, H. Qiu, S. Katti, M. Pavone, and S. Chinchali, "Personalized Federated Learning of Driver Prediction Models for Autonomous Driving," 2021, arXiv.
- [146] Z. Zhang, S. Wang, Y. Hong, L. Zhou, and Q. Hao, "Distributed Dynamic Map Fusion via Federated Learning for Intelligent Networked Vehicles," in 2021 IEEE International Conference on Robotics and Automation (ICRA), Xi'an, China: IEEE, May 2021, pp. 953–959.
- [147] M. Gao, L. Jin, Y. Jiang, and B. Guo, "Manifold Siamese Network: A Novel Visual Tracking ConvNet for Autonomous Vehicles," *IEEE Trans. Intell. Transport. Syst.*, vol. 21, no. 4, pp. 1612–1623, Apr. 2020.
- [148] W. Song, Y. Yang, M. Fu, F. Qiu, and M. Wang, "Real-Time Obstacles Detection and Status Classification for Collision Warning in

- a Vehicle Active Safety System," *IEEE Trans. Intell. Transport. Syst.*, vol. 19, no. 3, pp. 758–773, Mar. 2018.
- [149] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA: IEEE, Jun. 2016, pp. 779–788.
- [150] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation," in 2014 IEEE Conf. on Computer Vision and Pattern Recognition, Columbus, OH, USA: IEEE, Jun. 2014, pp. 580–587.
- [151] J. Huang et al., "Speed/Accuracy Trade-Offs for Modern Convolutional Object Detectors," in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI: IEEE, Jul. 2017, pp. 3296–3297.
- [152] X. Wu, X. Yao, and C.-L. Wang, "FedSCR: Structure-based Communication Reduction for Federated Learning," *IEEE Trans. Parallel Distrib. Syst.*, pp. 1–1, 2020.
- [153] D. Jallepalli, N. C. Ravikumar, P. V. Badarinath, S. Uchil, and M. A. Suresh, "Federated Learning for Object Detection in Autonomous Vehicles," in 2021 IEEE Seventh International Conference on Big Data Computing Service and Applications (BigDataService), Oxford, United Kingdom: IEEE, Aug. 2021, pp. 107–114.
- [154] G. Rjoub, O. A. Wahab, J. Bentahar, and A. S. Bataineh, "Improving Autonomous Vehicles Safety in Snow Weather Using Federated YOLO CNN Learning," in *Mobile Web and Intelligent Information* Systems, vol. 12814, J. Bentahar, I. Awan, M. Younas, and T.-M. Grønli, Eds., in Lecture Notes in Computer Science, vol. 12814., Cham: Springer International Publishing, 2021, pp. 121–134.
- [155] L. Barbieri, S. Savazzi, and M. Nicoli, "Decentralized Federated Learning for Road User Classification in Enhanced V2X Networks," in 2021 IEEE International Conference on Communications Workshops (ICC Workshops), Montreal, QC, Canada: IEEE, Jun. 2021, pp. 1–6.
- [156] S. Alshammari and S. Song, "3Pod: Federated Learning-based 3 Dimensional Pothole Detection for Smart Transportation," in 2022 IEEE International Smart Cities Conference (ISC2), Pafos, Cyprus: IEEE, Sep. 2022, pp. 1–7.
- [157] Y. Yuan, Y. Yuan, T. Baker, L. M. Kolbe, and D. Hogrefe, "FedRD: Privacy-preserving adaptive Federated learning framework for intelligent hazardous Road Damage detection and warning," *Future Generation Computer Systems*, vol. 125, pp. 385–398, Dec. 2021. future.2021.06.035.
- [158] I. V. Vondikakis, I. E. Panagiotopoulos, and G. J. Dimitrakopoulos, "FedRSC: A Federated Learning Analysis for Multi-Label Road Surface Classifications," *IEEE Open J. Intell. Transp. Syst.*, vol. 5, pp. 433–444, 2024.
- [159] P. K. Saha, D. Arya, and Y. Sekimoto, "Federated learning-based global road damage detection," *Computer aided Civil Eng*, vol. 39, no. 14, pp. 2223–2238, Jul. 2024.
- [160] X. Kong et al., "A Federated Learning-Based License Plate Recognition Scheme for 5G-Enabled Internet of Vehicles," IEEE Trans. Ind. Inf., vol. 17, no. 12, pp. 8523–8530, Dec. 2021.
- [161] K. Xie et al., "Efficient Federated Learning With Spike Neural Networks for Traffic Sign Recognition," *IEEE Trans. Veh. Technol.*, vol. 71, no. 9, pp. 9980–9992, Sep. 2022.
- [162] S. Wang et al., "Federated Deep Learning Meets Autonomous Vehicle Perception: Design and Verification," *IEEE Network*, vol. 37, no. 3, pp. 16–25, May 2023.
- [163] Z. Zhang et al., "Robust Semisupervised Federated Learning for Images Automatic Recognition in Internet of Drones," IEEE Internet Things J., vol. 10, no. 7, pp. 5733–5746, Apr. 2023.
- [164] D. Ye, R. Yu, M. Pan, and Z. Han, "Federated Learning in Vehicular Edge Computing: A Selective Model Aggregation Approach," *IEEE Access*, vol. 8, pp. 23920–23935, 2020.
- [165] H. Zhang, J. Bosch, H. H. Olsson, and A. C. Koppisetty, "AF-DNDF: Asynchronous Federated Learning of Deep Neural Decision Forests," in 2021 47th Euromicro Conference on Software Engineering and Advanced Applications (SEAA), Palermo, Italy: IEEE, Sep. 2021, pp. 308–315.
- [166] Y. Fu, F. R. Yu, C. Li, T. H. Luan, and Y. Zhang, "Vehicular Blockchain-Based Collective Learning for Connected and Autonomous Vehicles," *IEEE Wireless Commun.*, vol. 27, no. 2, pp. 197–203, Apr. 2020.

- [167] Y. Fu, C. Li, F. R. Yu, T. H. Luan, and Y. Zhang, "Hybrid Autonomous Driving Guidance Strategy Combining Deep Reinforcement Learning and Expert System," *IEEE Trans. Intell.* Transport. Syst., vol. 23, no. 8, pp. 11273–11286, Aug. 2022.
- [168] C. Li, Y. Fu, F. R. Yu, T. H. Luan, and Y. Zhang, "Vehicle Position Correction: A Vehicular Blockchain Networks-Based GPS Error Sharing Framework," *IEEE Trans. Intell. Transport. Syst.*, vol. 22, no. 2, pp. 898–912, Feb. 2021.
- [169] S. Otoum, I. Al Ridhawi, and H. T. Mouftah, "Blockchain-Supported Federated Learning for Trustworthy Vehicular Networks," in GLOBECOM 2020 - 2020 IEEE Global Communications Conference, Taipei, Taiwan: IEEE, Dec. 2020, pp. 1–6.
- [170] Y. Li, X. Tao, X. Zhang, J. Liu, and J. Xu, "Privacy-Preserved Federated Learning for Autonomous Driving," *IEEE Trans. Intell. Transport. Syst.*, vol. 23, no. 7, pp. 8423–8434, Jul. 2022.
- [171] P. Lv, L. Xie, J. Xu, X. Wu, and T. Li, "Misbehavior Detection in Vehicular Ad Hoc Networks Based on Privacy-Preserving Federated Learning and Blockchain," *IEEE Trans. Netw. Serv. Manage.*, vol. 19, no. 4, pp. 3936–3948, Dec. 2022.
- [172] R. Kumar, P. Kumar, R. Tripathi, G. P. Gupta, and N. Kumar, "P2SF-IoV: A Privacy-Preservation-Based Secured Framework for Internet of Vehicles," *IEEE Trans. Intell. Transport. Syst.*, vol. 23, no. 11, pp. 22571–22582, Nov. 2022.
- [173] H. Liu et al., "Blockchain and Federated Learning for Collaborative Intrusion Detection in Vehicular Edge Computing," *IEEE Trans. Veh. Technol.*, vol. 70, no. 6, pp. 6073–6084, Jun. 2021.
- [174] M. Abdel-Basset, N. Moustafa, H. Hawash, I. Razzak, K. M. Sallam, and O. M. Elkomy, "Federated Intrusion Detection in Blockchain-Based Smart Transportation Systems," *IEEE Trans. Intell. Transport. Syst.*, vol. 23, no. 3, pp. 2523–2537, Mar. 2022.
- [175] A. Boudguiga, O. Stan, A. Fazzat, H. Labiod, and P.-E. Clet, "Privacy Preserving Services for Intelligent Transportation Systems with Homomorphic Encryption:," in *Proceedings of the 7th International* Conference on Information Systems Security and Privacy, Online Streaming, --- Select a Country ---: SCITEPRESS - Science and Technology Publications, 2021, pp. 684–693.
- [176] C. Gentry, "Fully homomorphic encryption using ideal lattices," in Proceedings of the forty-first annual ACM symposium on Theory of computing, Bethesda MD USA: ACM, May 2009, pp. 169–178.
- [177] S. Sinaei, M. Mohammadi, R. Shrestha, M. Alibeigi, and D. Eklund, "PRIV-DRIVE: Privacy-Ensured Federated Learning using Homomorphic Encryption for Driver Fatigue Detection," in 2024 27th Euromicro Conference on Digital System Design (DSD), Paris, France: IEEE, Aug. 2024, pp. 427–434.
- [178] T. Li, L. Lin, and S. Gong, "AutoMPC: Efficient Multi-Party Computation for Secure and Privacy-Preserving Cooperative Control of Connected Autonomous Vehicles.," SafeAI@ AAAI, vol. 1, 2019.
- [179] K. Zhang, K. Chen, Z. Li, J. Chen, and Y. Zheng, "Privacy-Preserving Data-Enabled Predictive Leading Cruise Control in Mixed Traffic," 2022, arXiv.
- [180] K. Hangdong, M. Bo, H. Darong, and D. Zhaoyang, "FedVPS: Federated Learning for Privacy and Security of Internet of Vehicles on Non-IID Data," in 2023 IEEE 12th Data Driven Control and Learning Systems Conference (DDCLS), Xiangtan, China: IEEE, May 2023, pp. 178–183.
- [181] M. Liu, L. Cheng, Y. Gu, Y. Wang, Q. Liu, and N. E. O'Connor, "MPC-CSAS: Multi-Party Computation for Real-Time Privacy-Preserving Speed Advisory Systems," *IEEE Trans. Intell. Transport.* Syst., vol. 23, no. 6, pp. 5887–5893, Jun. 2022.
- [182] Y. Duan, J. Liu, X. Ming, W. Jin, Z. Song, and X. Peng, "Characterizing and Optimizing Differentially-Private Techniques for High-Utility, Privacy-Preserving Internet-of-Vehicles," in HCI in Mobility, Transport, and Automotive Systems, vol. 14048, H. Krömker, Ed., in Lecture Notes in Computer Science, vol. 14048., Cham: Springer Nature Switzerland, 2023, pp. 31–50.
- [183] K. Al-Hussaeni, B. C. M. Fung, F. Iqbal, G. G. Dagher, and E. G. Park, "SafePath: Differentially-private publishing of passenger trajectories in transportation systems," *Computer Networks*, vol. 143, pp. 126–139, Oct. 2018.
- [184] Z. Zhou, Y. Qiao, L. Zhu, J. Guan, Y. Liu, and C. Xu, "Differential privacy-guaranteed trajectory community identification over vehicle ad-hoc networks," *Internet Technology Letters*, vol. 1, no. 3, p. e9, May 2018.

- [185] Z. Ma, T. Zhang, X. Liu, X. Li, and K. Ren, "Real-Time Privacy-Preserving Data Release Over Vehicle Trajectory," *IEEE Trans. Veh. Technol.*, vol. 68, no. 8, pp. 8091–8102, Aug. 2019.
- [186] T. Zhang and Q. Zhu, "Distributed Privacy-Preserving Collaborative Intrusion Detection Systems for VANETs," *IEEE Trans. on Signal and Inf. Process. over Networks*, vol. 4, no. 1, pp. 148–161, Mar. 2018
- [187] G. Raja, S. Anbalagan, G. Vijayaraghavan, S. Theerthagiri, S. V. Suryanarayan, and X.-W. Wu, "SP-CIDS: Secure and Private Collaborative IDS for VANETs," *IEEE Trans. Intell. Transport. Syst.*, vol. 22, no. 7, pp. 4385–4393, Jul. 2021.
- [188] Q. Xu, L. Zhang, D. Ou, and W. Yu, "Secure Intrusion Detection by Differentially Private Federated Learning for Inter-Vehicle Networks," *Transportation Research Record: J. of the Transportation Research Board*, vol. 2677, no. 9, pp. 421–437, Sep. 2023.
- [189] Z. Iftikhar, A. Anjum, A. Khan, M. A. Shah, and G. Jeon, "Privacy preservation in the internet of vehicles using local differential privacy and IOTA ledger," *Cluster Comput*, vol. 26, no. 6, pp. 3361–3377, Dec. 2023.
- [190] Y. Lu, X. Huang, Y. Dai, S. Maharjan, and Y. Zhang, "Federated Learning for Data Privacy Preservation in Vehicular Cyber-Physical Systems," *IEEE Network*, vol. 34, no. 3, pp. 50–56, May 2020.
- [191] Y. Lei, S. L. Wang, C. Su, and T. F. Ng, "OES-Fed: a federated learning framework in vehicular network based on noise data filtering," *PeerJ Computer Science*, vol. 8, p. e1101, Sep. 2022.
- [192] P. Zheng, Y. Zhu, Y. Hu, and A. Schmeink, "Data-driven Extreme Events Modeling for Vehicle Networks by Personalized Federated Learning: Invited Paper," in 2022 International Symposium on Wireless Communication Systems (ISWCS), Hangzhou, China: IEEE, Oct. 2022, pp. 1–6.
- [193] X. Li, L. Cheng, C. Sun, K.-Y. Lam, X. Wang, and F. Li, "Federated-Learning-Empowered Collaborative Data Sharing for Vehicular Edge Networks," *IEEE Network*, vol. 35, no. 3, pp. 116–124, May 2021.
- [194] F. O. Olowononi, D. B. Rawat, and C. Liu, "Federated Learning with Differential Privacy for Resilient Vehicular Cyber Physical Systems," in 2021 IEEE 18th Annual Consumer Communications & Networking Conference (CCNC), Las Vegas, NV, USA: IEEE, Jan. 2021, pp. 1–5.
- [195] F. Ayaz, Z. Sheng, D. Tian, and Y. L. Guan, "A Blockchain Based Federated Learning for Message Dissemination in Vehicular Networks," *IEEE Trans. Veh. Technol.*, vol. 71, no. 2, pp. 1927–1940, Feb. 2022.
- [196] X. Huang, P. Li, R. Yu, Y. Wu, K. Xie, and S. Xie, "FedParking: A Federated Learning Based Parking Space Estimation With Parked Vehicle Assisted Edge Computing," *IEEE Trans. Veh. Technol.*, vol. 70, no. 9, pp. 9355–9368, Sep. 2021.
- [197] Y. Wu, Z. Shen, Y. Tian, Z. Cai, and F. Li, "Electric vehicle charging load forecasting based on federal learning," in *International Conference on Electronic Information Engineering and Computer Communication (EIECC 2021)*, Z. Zhu and F. Cen, Eds., Nanchang, China: SPIE, May 2022, p. 56.
- [198] Y. M. Saputra, D. T. Hoang, D. N. Nguyen, E. Dutkiewicz, M. D. Mueck, and S. Srikanteswara, "Energy Demand Prediction with Federated Learning for Electric Vehicle Networks," in 2019 IEEE Global Communications Conference (GLOBECOM), Waikoloa, HI, USA: IEEE, Dec. 2019, pp. 1–6.
- [199] X. Wang, X. Zheng, and X. Liang, "Charging Station Recommendation for Electric Vehicle Based on Federated Learning," J. Phys.: Conf. Ser., vol. 1792, no. 1, p. 012055, Feb. 2021.
- [200] Y. M. Saputra, D. N. Nguyen, D. T. Hoang, T. X. Vu, E. Dutkiewicz, and S. Chatzinotas, "Federated Learning Meets Contract Theory: Economic-Efficiency Framework for Electric Vehicle Networks," *IEEE Trans. on Mobile Computing*, vol. 21, no. 8, pp. 2803–2817, Aug. 2022.
- [201] A. T. Thorgeirsson, S. Scheubner, S. Funfgeld, and F. Gauterin, "Probabilistic Prediction of Energy Demand and Driving Range for Electric Vehicles With Federated Learning," *IEEE Open J. Veh. Technol.*, vol. 2, pp. 151–161, 2021.
- [202] M. Wilbur, C. Samal, J. P. Talusan, K. Yasumoto, and A. Dubey, "Time-dependent Decentralized Routing using Federated Learning," in 2020 IEEE 23rd International Symposium on Real-Time Distributed Computing (ISORC), Nashville, TN, USA: IEEE, May 2020, pp. 56–64.

- [203] T. Zeng et al., "Multi-Task Federated Learning for Traffic Prediction and Its Application to Route Planning," in 2021 IEEE Intell. Vehicles Symposium (IV), Nagoya, Japan: IEEE, Jul. 2021, pp. 451–457.
- [204] Z. Zhang, H. Wang, Z. Fan, J. Chen, X. Song, and R. Shibasaki, "GOF-TTE: Generative Online Federated Learning Framework for Travel Time Estimation," *IEEE Internet Things J.*, vol. 9, no. 23, pp. 24107–24121, Dec. 2022.
- [205] Y. Zhu, Y. Ye, Y. Liu, and J. J. Q. Yu, "Cross-Area Travel Time Uncertainty Estimation From Trajectory Data: A Federated Learning Approach," *IEEE Transaction on Intelligent Transport. Syst.*, vol. 23, no. 12, pp. 24966–24978, Dec. 2022.
- [206] Q. Liu, L. Shi, L. Sun, J. Li, M. Ding, and F. S. Shu, "Path Planning for UAV-Mounted Mobile Edge Computing With Deep Reinforcement Learning," *IEEE Trans. Veh. Technol.*, vol. 69, no. 5, pp. 5723–5728, May 2020.
- [207] A. Shahbazi, I. Donevski, J. J. Nielsen, and M. Di Renzo, "Federated Reinforcement Learning UAV Trajectory Design for Fast Localization of Ground Users," in 2022 30th European Signal Processing Conf. (EUSIPCO), Belgrade, Serbia: IEEE, Aug. 2022, pp. 663–666.
- [208] J. Konečný, H. B. McMahan, F. X. Yu, P. Richtárik, A. T. Suresh, and D. Bacon, "Federated Learning: Strategies for Improving Communication Efficiency," 2016, arXiv.
- [209] R. Song, L. Lyu, W. Jiang, A. Festag, and A. Knoll, "V2X-Boosted Federated Learning for Cooperative Intelligent Transportation Systems with Contextual Client Selection," 2023, arXiv.
- [210] S. B. Prathiba, G. Raja, S. Anbalagan, K. Dev, S. Gurumoorthy, and A. P. Sankaran, "Federated Learning Empowered Computation Offloading and Resource Management in 6G-V2X," *IEEE Trans. Netw. Sci. Eng.*, vol. 9, no. 5, pp. 3234–3243, Sep. 2022.
- [211] X. Li, L. Lu, W. Ni, A. Jamalipour, D. Zhang, and H. Du, "Federated Multi-Agent Deep Reinforcement Learning for Resource Allocation of Vehicle-to-Vehicle Communications," *IEEE Trans. Veh. Technol.*, vol. 71, no. 8, pp. 8810–8824, Aug. 2022.
- [212] X. Zhou, W. Liang, J. She, Z. Yan, and K. Wang, "Two-Layer Federated Learning With Heterogeneous Model Aggregation for 6G Supported Internet of Vehicles," *IEEE Trans. Veh. Technol.*, vol. 70, no. 6, pp. 5308–5317, Jun. 2021.
- [213] S. Samarakoon, M. Bennis, W. Saad, and M. Debbah, "Distributed Federated Learning for Ultra-Reliable Low-Latency Vehicular Communications," *IEEE Trans. Commun.*, vol. 68, no. 2, pp. 1146– 1159, Feb. 2020.
- [214] Y. Wang, Y. Tian, X. Hei, L. Zhu, and W. Ji, "A Novel IoV Block-Streaming Service Awareness and Trusted Verification Scheme in 6G," *IEEE Transaction on Vehicle. Technology*, vol. 70, no. 6, pp. 5197–5210, Jun. 2021.
- [215] A. Shukla, P. Bhattacharya, S. Tanwar, N. Kumar, and M. Guizani, "DwaRa: A Deep Learning-Based Dynamic Toll Pricing Scheme for Intelligent Transportation Systems," *IEEE Trans. Veh. Technol.*, vol. 69, no. 11, pp. 12510–12520, Nov. 2020.
- [216] Z. Wan, T. Zhang, W. Liu, M. Wang, and L. Zhu, "Decentralized Privacy-Preserving Fair Exchange Scheme for V2G Based on Blockchain," *IEEE Trans. Dependable and Secure Comput.*, vol. 19, no. 4, pp. 2442–2456, Jul. 2022.
- [217] J. Gao et al., "A Blockchain-SDN-Enabled Internet of Vehicles Environment for Fog Computing and 5G Networks," *IEEE Internet Things J.*, vol. 7, no. 5, pp. 4278–4291, May 2020.
- [218] Y. Ren, X. Chen, S. Guo, S. Guo, and A. Xiong, "Blockchain-Based VEC Network Trust Management: A DRL Algorithm for Vehicular Service Offloading and Migration," *IEEE Trans. Veh. Technol.*, vol. 70, no. 8, pp. 8148–8160, Aug. 2021.
- [219] Y. Lu, X. Huang, K. Zhang, S. Maharjan, and Y. Zhang, "Blockchain Empowered Asynchronous Federated Learning for Secure Data Sharing in Internet of Vehicles," *IEEE Trans. Veh. Technol.*, vol. 69, no. 4, pp. 4298–4311, Apr. 2020.
- [220] H. Chai, S. Leng, Y. Chen, and K. Zhang, "A Hierarchical Blockchain-Enabled Federated Learning Algorithm for Knowledge Sharing in Internet of Vehicles," *IEEE Trans. Intell. Transport. Syst.*, vol. 22, no. 7, pp. 3975–3986, Jul. 2021.
- [221] M. Shen, J. Zhang, L. Zhu, K. Xu, and X. Tang, "Secure SVM Training Over Vertically-Partitioned Datasets Using Consortium Blockchain for Vehicular Social Networks," *IEEE Trans. Veh. Technol.*, vol. 69, no. 6, pp. 5773–5783, Jun. 2020.

- [222] S. R. Pokhrel and J. Choi, "Federated Learning With Blockchain for Autonomous Vehicles: Analysis and Design Challenges," *IEEE Trans. Commun.*, vol. 68, no. 8, pp. 4734–4746, Aug. 2020.
- [223] F. Yucel, K. Akkaya, and E. Bulut, "Efficient and privacy preserving supplier matching for electric vehicle charging," Ad Hoc Networks, vol. 90, p. 101730, Jul. 2019.
- [224] L. Chen, J. Zhou, Y. Chen, Z. Cao, X. Dong, and K.-K. R. Choo, "PADP: Efficient Privacy-Preserving Data Aggregation and Dynamic Pricing for Vehicle-to-Grid Networks," *IEEE Internet Things J.*, vol. 8, no. 10, pp. 7863–7873, May 2021.
- [225] B. Li, Y. Guo, Q. Du, Z. Zhu, X. Li, and R. Lu, "\$\bm {P}^{\bm {3}}\\$: Privacy-Preserving Prediction of Real-Time Energy Demands in EV Charging Networks," *IEEE Trans. Ind. Inf.*, vol. 19, no. 3, pp. 3029–3038, Mar. 2023.
- [226] F. Farokhi, I. Shames, and K. H. Johansson, "Private routing and ridesharing using homomorphic encryption," *IET cyber-phys. syst.*, vol. 5, no. 4, pp. 311–320, Dec. 2020.
- [227] H. Karim and D. B. Rawat, "TollsOnly Please—Homomorphic Encryption for Toll Transponder Privacy in Internet of Vehicles," *IEEE Internet Things J.*, vol. 9, no. 4, pp. 2627–2636, Feb. 2022.
- [228] A. Lakhan, T.-M. Groenli, H. Wu, M. Younas, and G. Ghinea, "A Novel Homomorphic Blockchain Scheme for Intelligent Transport Services in Fog/Cloud and IoT Networks," *IEEE Trans. Intell. Transport. Syst.*, pp. 1–16, 2024.
- [229] X. Guo, B. Wang, Y. Jiang, D. Zhang, and L. Cao, "Homomorphic encryption based privacy-aware intelligent forwarding mechanism for NDN-VANET," *ComSIS*, vol. 20, no. 1, pp. 1–24, 2023.
- [230] W. Jiang, J. Tao, and Z. Guan, "A Trusted Data Privacy Computing Method for Vehicular Ad Hoc Networks Based on Homomorphic Encryption and DAG Blockchain," *IEEE Internet Things Journal*, pp. 1–1, 2024.
- [231] G. Raja, Y. Manaswini, G. D. Vivekanandan, H. Sampath, K. Dev, and A. K. Bashir, "AI-Powered Blockchain A Decentralized Secure Multiparty Computation Protocol for IoV," in *IEEE INFOCOM 2020 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, Toronto, ON, Canada: IEEE, Jul. 2020, pp. 865–870.
- [232] Y. AlSaqabi and B. Krishnamachari, "Incentivizing Private Data Sharing in Vehicular Networks: A Game-Theoretic Approach," in 2023 IEEE 98th Vehicular Technology Conference (VTC2023-Fall), Hong Kong, Hong Kong: IEEE, Oct. 2023, pp. 1–8.
- [233] J. A. Khan, W. Wang, and K. Ozbay, "BELIEVE: Privacy-Aware Secure Multi-Party Computation for Real-Time Connected and Autonomous Vehicles and Micro-Mobility Data Validation Using Blockchain—A Study on New York City Data," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2678, no. 3, pp. 410–421, Mar. 2024.
- [234] H. Yang, J. Shen, T. Zhou, S. Ji, and P. Vijayakumar, "A Flexible and Privacy-Preserving Collaborative Filtering Scheme in Cloud Computing for VANETs," ACM Trans. Internet Technol., vol. 22, no. 2, pp. 1–19, May 2022.
- [235] R. Qiu, X. Liu, R. Huang, F. Zheng, L. Liang, and Y. Li, "Differential privacy EV charging data release based on variable window," *PeerJ Computer Science*, vol. 7, p. e481, Apr. 2021.
- [236] U. I. Atmaca, S. Biswas, C. Maple, and C. Palamidessi, "A Privacy-Preserving Querying Mechanism with High Utility for Electric Vehicles," *IEEE Open J. Veh. Technol.*, vol. 5, pp. 262–277, 2024.
- [237] A. A. Khaliq, A. Anjum, A. B. Ajmal, J. L. Webber, A. Mehbodniya, and S. Khan, "A Secure and Privacy Preserved Parking Recommender System Using Elliptic Curve Cryptography and Local Differential Privacy," *IEEE Access*, vol. 10, pp. 56410–56426, 2022.
- [238] H. Batool, A. Anjum, A. Khan, S. Izzo, C. Mazzocca, and G. Jeon, "A secure and privacy preserved infrastructure for VANETs based on federated learning with local differential privacy," *Information Sciences*, vol. 652, p. 119717, Jan. 2024.
- [239] Y. Zhao, M. Li, L. Lai, N. Suda, D. Civin, and V. Chandra, "Federated Learning with Non-IID Data," 2018.
- [240] M. Fredrikson, S. Jha, and T. Ristenpart, "Model Inversion Attacks that Exploit Confidence Information and Basic Countermeasures," in Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security, Denver Colorado USA: ACM, Oct. 2015, pp. 1322–1333.
- [241] M. Nasr, R. Shokri, and A. Houmansadr, "Comprehensive Privacy Analysis of Deep Learning: Passive and Active White-box Inference

- Attacks against Centralized and Federated Learning," in 2019 IEEE Symposium on Security and Privacy (SP), San Francisco, CA, USA: IEEE, May 2019, pp. 739–753.
- [242] K. Pang, T. Qi, C. Wu, M. Bai, M. Jiang, and Y. Huang, "ModelShield: Adaptive and Robust Watermark against Model Extraction Attack," *IEEE Transactions on Information Forensics and Security*, pp. 1–1, 2025.
- [243] J. Yang, J. Zheng, T. Baker, S. Tang, Y. Tan, and Q. Zhang, "Clean-label poisoning attacks on federated learning for IoT," Expert Systems, vol. 40, no. 5, p. e13161, Jun. 2023.
- [244] P. Blanchard, E. M. El Mhamdi, R. Guerraoui, and J. Stainer, "Machine learning with adversaries: Byzantine tolerant gradient descent," Advances in neural info. processing systems, vol. 30, 2017.
- [245] D. Yin, Y. Chen, R. Kannan, and P. Bartlett, "Byzantine-robust distributed learning: Towards optimal statistical rates," in *International conf. on machine learning*, Pmlr, 2018, pp. 5650–5659.
- [246] X. Ma and D. Xu, "TORR: A Lightweight Blockchain for Decentralized Federated Learning," *IEEE Internet Things J.*, vol. 11, no. 1, pp. 1028–1040, Jan. 2024.
- [247] V. Costan, "Intel SGX explained," IACR Cryptol, EPrint Arch, 2016.
- [248] A. Kim, A. Papadimitriou, and Y. Polyakov, "Approximate Homomorphic Encryption with Reduced Approximation Error," in Topics in Cryptology - CT-RSA 2022, vol. 13161, S. D. Galbraith, Ed., in Lecture Notes in Computer Science, vol. 13161. , Cham: Springer International Publishing, 2022, pp. 120–144.
- [249] C. Silvano et al., "A Survey on Deep Learning Hardware Accelerators for Heterogeneous HPC Platforms," 2023, arXiv.
- [250] B. Jayaraman and D. Evans, "Evaluating differentially private machine learning in practice," in 28th USENIX Security Symposium (USENIX Security 19), 2019, pp. 1895–1912.
- [251] B. Kulynych, J. F. Gomez, G. Kaissis, F. du P. Calmon, and C. Troncoso, "Attack-Aware Noise Calibration for Differential Privacy," 2024, arXiv.
- [252] P. Guerra-Balboa, A. Miranda-Pascual, J. Parra-Arnau, and T. Strufe, "Composition in Differential Privacy for General Granularity Notions (Long Version)," 2023, arXiv.
- [253] M. Bun and T. Steinke, "Concentrated Differential Privacy: Simplifications, Extensions, and Lower Bounds," in *Theory of Cryptography*, vol. 9985, M. Hirt and A. Smith, Eds., in Lecture Notes in Computer Science, vol. 9985, Berlin, Heidelberg: Springer Berlin Heidelberg, 2016, pp. 635–658.
- [254] S. Pinto and N. Santos, "Demystifying Arm TrustZone: A Comprehensive Survey," ACM Comput. Surv., vol. 51, no. 6, pp. 1– 36, Nov. 2019.
- [255] S. Han, H. Mao, and W. J. Dally, "Deep Compression: Compressing Deep Neural Networks with Pruning, Trained Quantization and Huffman Coding," 2015, arXiv.
- [256] G. Xie et al., "Sequential Trajectory Data Publishing With Adaptive Grid-Based Weighted Differential Privacy," *IEEE Trans. Knowl. Data Eng.*, vol. 36, no. 12, pp. 9249–9262, Dec. 2024.
- [257] L. Ducas *et al.*, "CRYSTALS-Dilithium: A Lattice-Based Digital Signature Scheme," *TCHES*, pp. 238–268, Feb. 2018.
- [258] N. Islam and S. Shin, "Robust Deep Learning Models for OFDM-Based Image Communication Systems in Intelligent Transportation Systems (ITS) for Smart Cities," *Electronics*, vol. 12, no. 11, p. 2425, May 2023.
- [259] C. Szegedy et al., "Intriguing properties of neural networks," 2013, arXiv.
- [260] P. Iyenghar, E. Gracic, and G. Pawelke, "A Systematic Approach to Enhancing ISO 26262 With Machine Learning-Specific Life Cycle Phases and Testing Methods," *IEEE Access*, vol. 12, pp. 179600– 179627, 2024.
- [261] X. Geng et al., "From Algorithm to Hardware: A Survey on Efficient and Safe Deployment of Deep Neural Networks," IEEE Trans. Neural Netw. Learning Syst., pp. 1–21, 2024.
- [262] Z. Niu, G. Zhong, and H. Yu, "A review on the attention mechanism of deep learning," *Neurocomputing*, vol. 452, pp. 48–62, Sep. 2021.
- [263] D. Gunning, M. Stefik, J. Choi, T. Miller, S. Stumpf, and G.-Z. Yang, "XAI—Explainable artificial intelligence," *Sci. Robot.*, vol. 4, no. 37, p. eaay7120, Dec. 2019.
- [264] C. Jin, X. Feng, and H. Yu, "Embracing Multiheterogeneity and Privacy Security Simultaneously: A Dynamic Privacy-Aware Federated Reinforcement Learning Approach," *IEEE Trans. Neural Netw. Learning Syst.*, pp. 1–15, 2024.

- [265] M. Ul Hassan, M. H. Rehmani, M. Rehan, and J. Chen, "Differential Privacy in Cognitive Radio Networks: A Comprehensive Survey," *Cogn Comput*, vol. 14, no. 2, pp. 475–510, Mar. 2022.
- [266] R. Pal, A. Prakash, R. Tripathi, and K. Naik, "Regional Super Cluster Based Optimum Channel Selection for CR-VANET," *IEEE Trans. Cogn. Commun. Netw.*, vol. 6, no. 2, pp. 607–617, Jun. 2020.
- [267] T. Chen, H. Lu, T. Kunpittaya, and A. Luo, "A Review of zk-SNARKs," 2022, arXiv.
- [268] I. Chillotti, N. Gama, M. Georgieva, and M. Izabachène, "TFHE: Fast Fully Homomorphic Encryption Over the Torus," *J Cryptol*, vol. 33, no. 1, pp. 34–91, Jan. 2020.