IEEE
ITSC 2023
BILBAO

# *A Survey of Federated Learning for Connected and Automated Vehicles*


Vishnu Pandi Chellapandi

Liangqi Yuan

Stanislaw H Żak

Ziran Wang


September 28, 2023


PURDUE UNIVERSITY


This Paper arXiv

Journal Extension arXiv

# *Motivation for Federated Learning for CAVs*

- CAVs generate massive amounts of raw data, between 20 and 40 TB per day, per vehicle from various sources such as engine components, electronic control units (ECU), perception sensors, and vehicle-to-everything (V2X) communications

- Not feasible to have a secure framework to collect this large amount of data from every vehicle and train an ML model. This led to the development of a new ML paradigm known as **Federated Learning (FL)**

- In FL, edge devices/clients only send the learnable parameters to cloud servers rather than sending massive local datasets in a centralized learning framework

- Cloud servers perform secure aggregation of the received parameters and update the global model parameters that are transmitted back to the vehicles

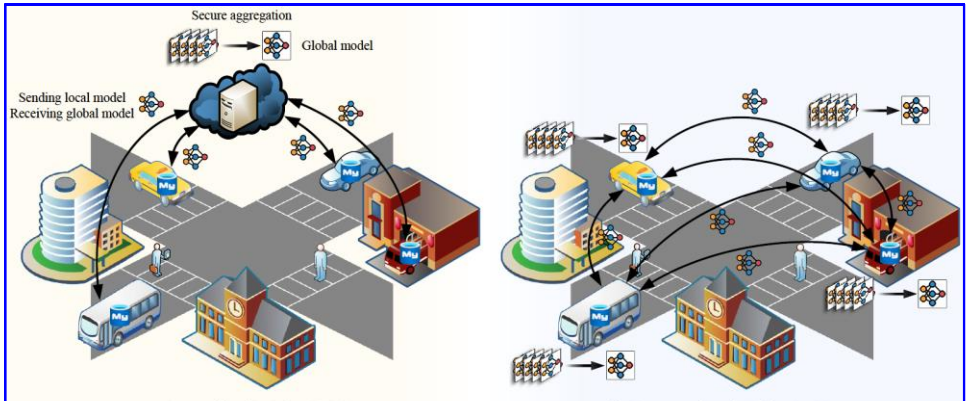
# *Centralized vs Decentralized FL*

COMPARISON OF MACHINE LEARNING APPROACHES IN CONNECTED AND AUTOMATED VEHICLES

| **Features** | **Edge Learning (On-Vehicle only)** | **Centralized Learning (On-Server only)** | **Centralized Federated Learning** | **Decentralized Federated Learning** |
|---|---|---|---|---|
| Model training | Local vehicle | Central server | Local vehicle training and central server aggregation | Local vehicle training and aggregation |
| Model applicability | Personalized model | Single global model | Single global model but can be personalized | Global models and personalized models |
| Privacy protection | ✓✓ | ✗ | ✓ | ▲ |
| Learning efficiency | ▲ | ✓ | ✓✓ | ✓ |
| Performance on heterogeneous/anomaly data | ▲ | ✓✓ | ✓ | ✓✓ |
| Communication (Data transmission) requirement | ✓✓ | ✗ | ▲ | ✗ |
| Training data volume | ✗ | ✓✓ | ✓ | ✓ |
| Current research progress | ✓✓ | ✓✓ | ▲ | ✗ |
| Compatibility with CAV | ✓ | ✗ | ✓ | ✓✓ |

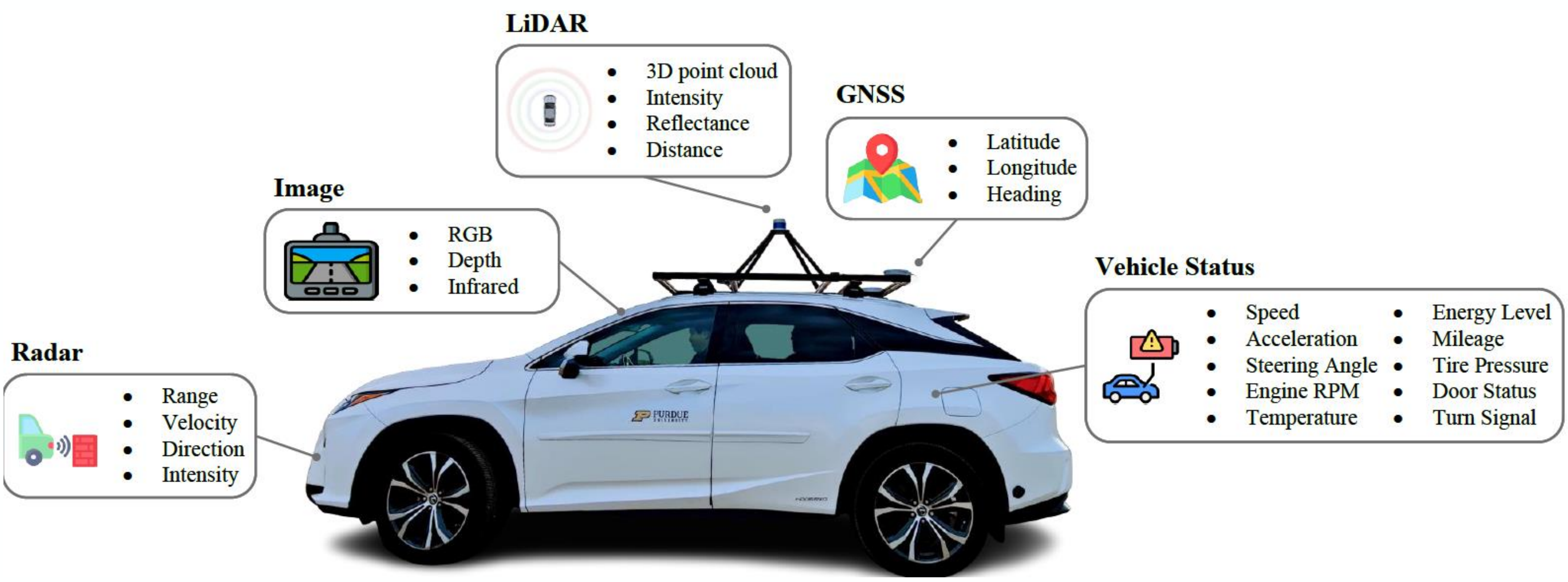
✓✓ Very high, ✓ high, ▲ average, ✗ low.

- In the **CFL** paradigm, model parameters are transmitted to a central server for aggregation
- **DFL** relies on a consensus among the vehicles, fostering collaboration to collectively update global parameters without the need for a central server




(a) Centralized FL for CAVs (b) Decentralized FL for CAVs

# Data Modalities




PURDUE UNIVERSITY

# *Applications – In-Vehicle Human Monitoring*

- Ability to enhance driver safety and personalization of assistance systems while preserving individual privacy by training ML models locally on each vehicle's data without sharing sensitive information centrally

(a) StateFarm

(b) AICity

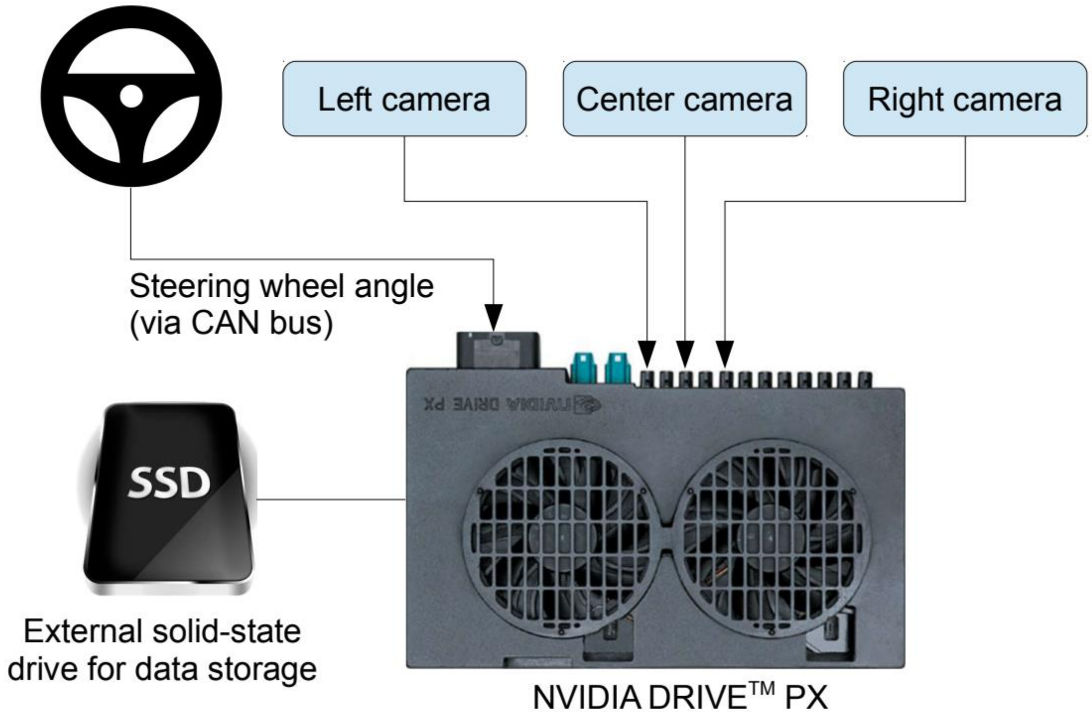


Datasets u

Yuan *et al.* (2023)

# *Applications – Steering Wheel Angle Prediction*

- Ability to collectively train accurate and personalized models across a fleet of vehicles without sharing sensitive data, enhancing overall safety and performance

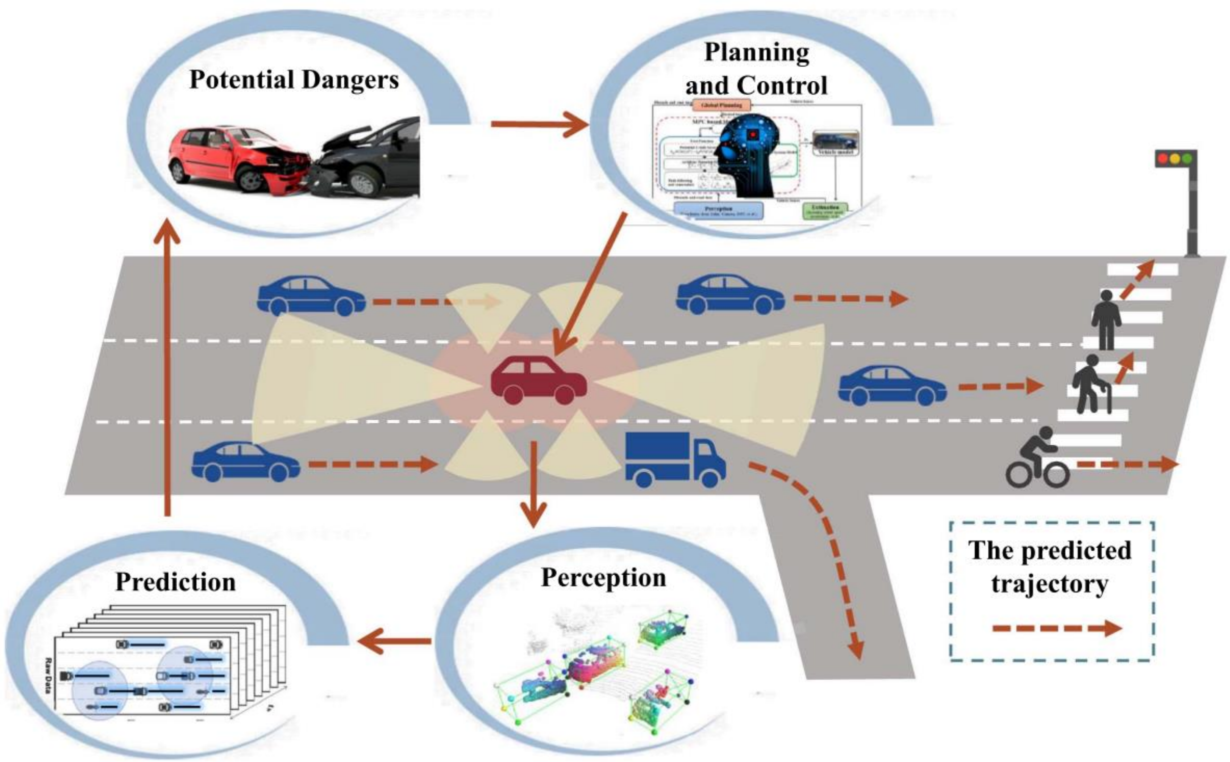





Bojarski *et al.* (2016)

# *Applications - Vehicle Trajectory Prediction*

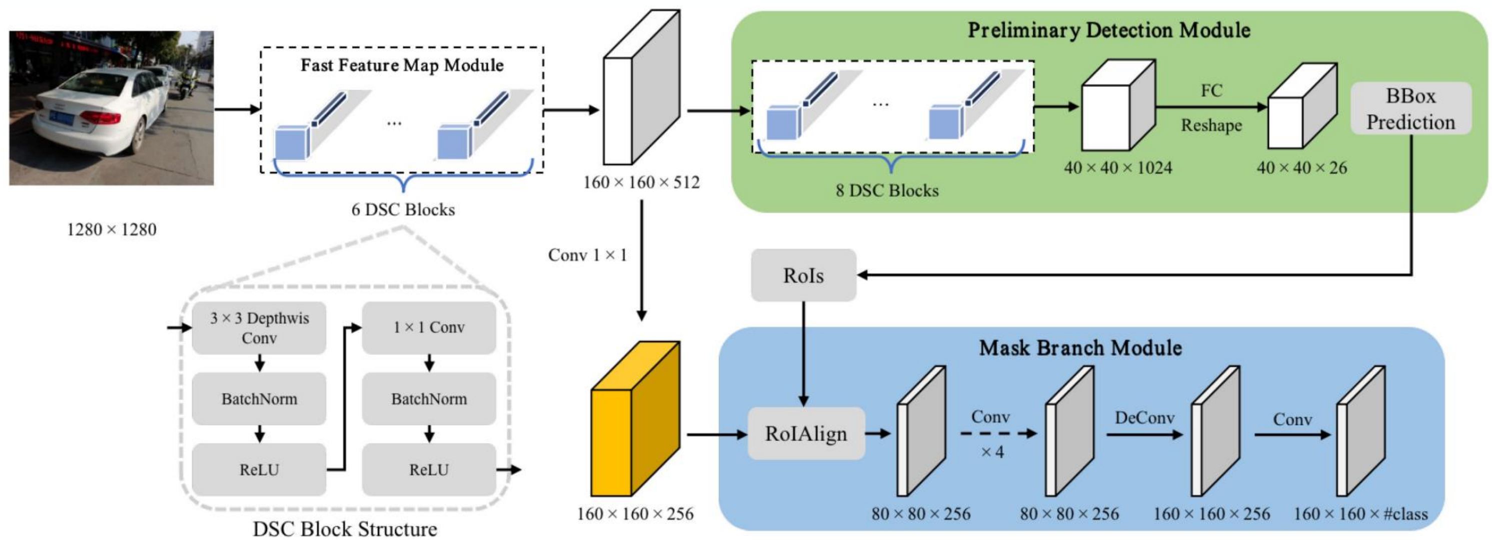
- Capacity to collaboratively improve prediction accuracy and enhances generalizability across a network of vehicles while maintaining data privacy and security, leading to safer and more efficient transportation systems




Huang *et al.* (2022)

# Applications - Object Detection in Vehicles

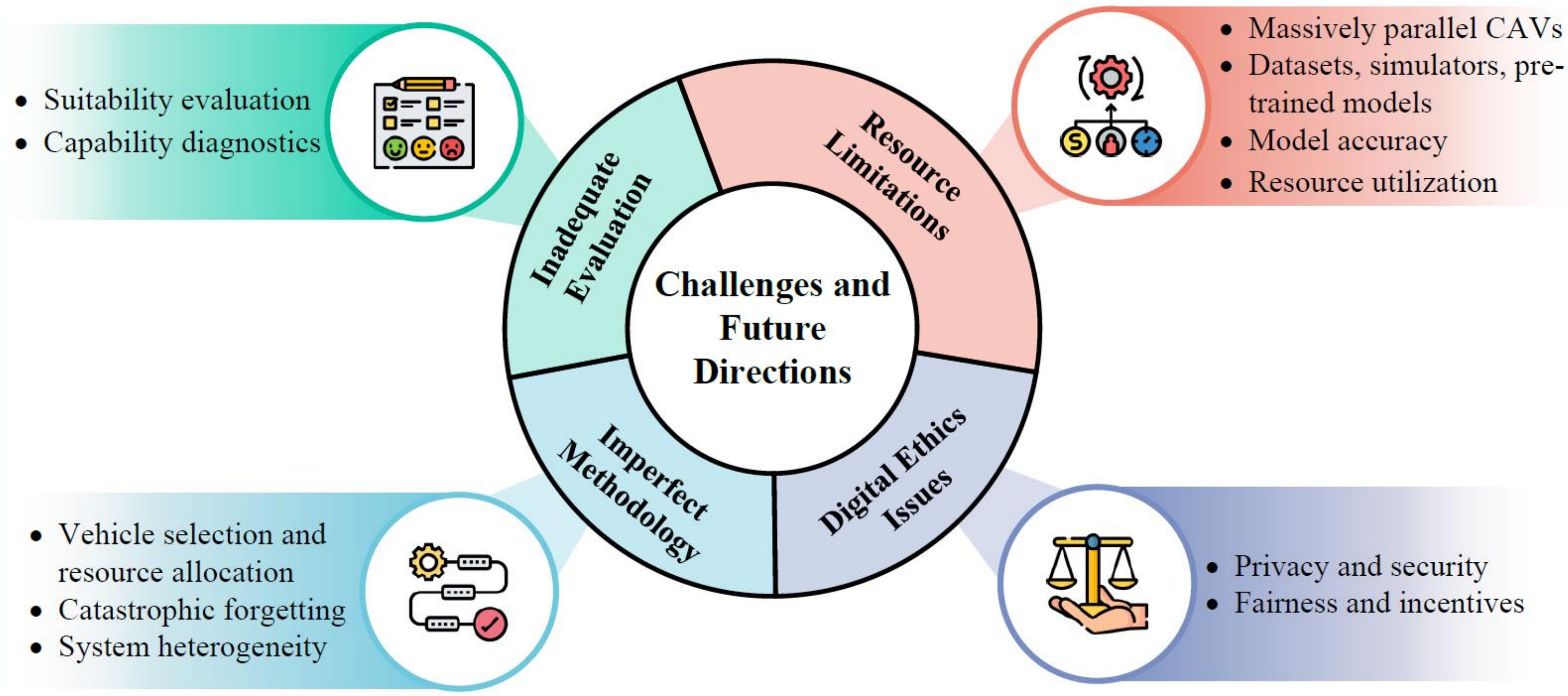
- Collectively train and improve detection models using decentralized data from multiple vehicles, enhancing learning, safety and accuracy while preserving data privacy




PURDUE UNIVERSITY

Kong *et al.* (2021)

# Challenges of FL4CAV




PURDUE UNIVERSITY

# *Thank You*

***Contact Us!***


| | |
|---|---|
| Vishnu Pandi Chellapandi | cvp@purdue.edu |
| Liangqi Yuan | liangqiy@purdue.edu |
| Stanislaw H Żak | zak@purdue.edu |
| Ziran Wang | ziran@purdue.edu |


PURDUE UNIVERSITY®