DENSO
Crafting the Core

Challenges of Big Data, Artificial Intelligence and Vehicle Data

July 2022
Christian Prehofer

Big Data for Vehicle Data Analysis

- **Big Data for connected vehicle applications**
  - Enormous amount of data
  - Many applications
- **Use case driving behavior & energy efficiency**
  - Compute efficiency for every second
  - **Comparison** of Big Data processing options
- **Use case driver status monitoring**
  - Privacy preserving data analysis with federated learning
- **Discussion and Outlook**

Motivation – Big Data and Data Analysis in Automotive

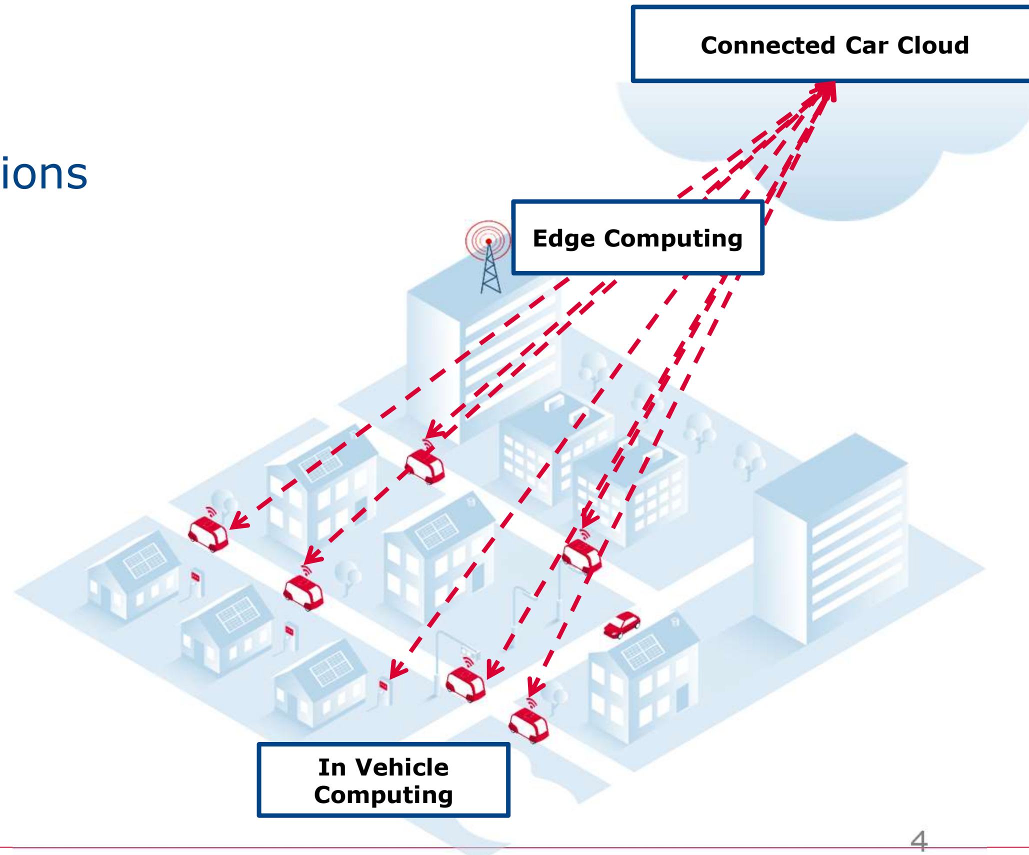
Vehicle Driving Data Applications:
e.g. insurance, eco driving, predictive maintenance, ADAS / Autonomous Driving

2TB/day from internal CAN bus

Vehicle Sensor Data: 2 TB/hour

Applications of Connected Vehicles

- **Enhancing in-vehicle functions**
  - Routing and traffic data
  - Energy efficient driving
  - Enhanced autonomous driving functions
- **New services**
  - Insurance based on actual driving
  - Car sharing
  - In-car payment (fuel, …)
- **Management**
  - Predictive maintenance
  - SW / function updates



Source: pixabay

Christian Prehofer, Denso Europe R&D

Big Data for Vehicle Data Analysis

- **Big Data for connected vehicle applications**
  - Enormous amount of data
  - Many applications

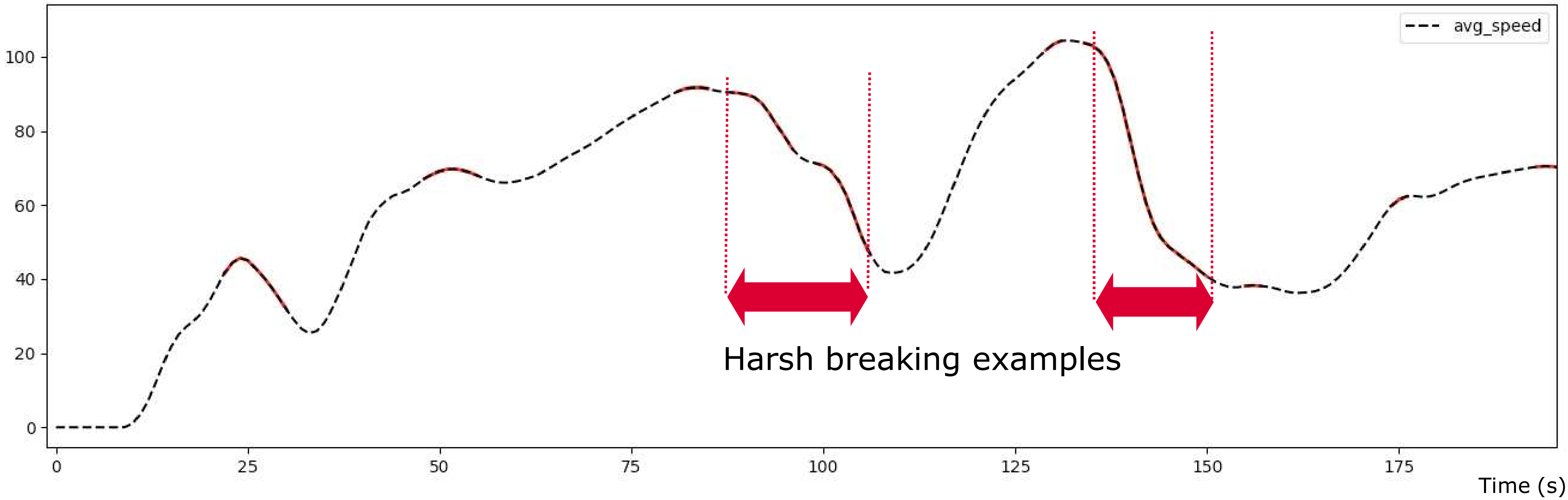
- **Use case driving behavior & energy efficiency**
  - Compute efficiency for every second
  - **Comparison** of Big Data processing options

- **Use case driver status monitoring**
  - Privacy preserving data analysis with federated learning

- **Discussion and Outlook**

Example Driving Behavior: Harsh breaking

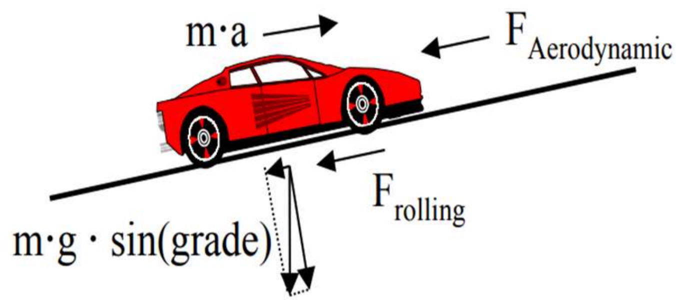
- Find out breaking phases based on speed and acceleration
- Hard brake: deceleration is greater than a certain threshold
- Example trip:



Use Case: Energy Efficiency

- Public **data set** (>500 trips, 8000km), incl.
  - Location
  - Speed
  - Energy consumption
  - Air conditioning, heating
  - Vehicle information (weight), …
- Calculate **„needed energy“**
  - **VSP: Vehicle specific power**
  - Need road inclination (from GPS coordinates), acceleration etc

https://github.com/gsoh/VED/blob/master/README.md



$$VSP\left[\frac{W}{kg}\right] = \frac{Power}{Mass} = \frac{\frac{d}{dt}(E_{kinetic} + E_{potential}) + F_{rolling} \times v - F_{aerodynamic} \times v}{m}$$

Use Case: Comparing Used and Needed Energy

Driving data for e-vehicles
- Speed
- Uphill/downhill
- Vehicle weight

Calculate **needed energy: VSP**
(Vehicle Specific Power, physical model)

$$VSP\left[\frac{W}{kg}\right] = \frac{Power}{Mass} = \frac{\frac{d}{dt}(E_{kinetic} + E_{potential}) + F_{rolling} \times v - F_{aerodynamic} \times v}{m}$$

Simplified version

$$VSP \approx v \cdot [a \cdot 1.1 + 9.81 \cdot grade + 0.213\ ] + 0.000426 \cdot v^3$$

Vehicle **energy consumption**
- KWh from e-vehicle data
- Consider AC and heating
- Temperature, Battery SOC

Needed vs actual energy
- Energy efficiency calculation
  - different driving phases
- Energy in different temperatures

Use Case: Energy Efficiency Analysis

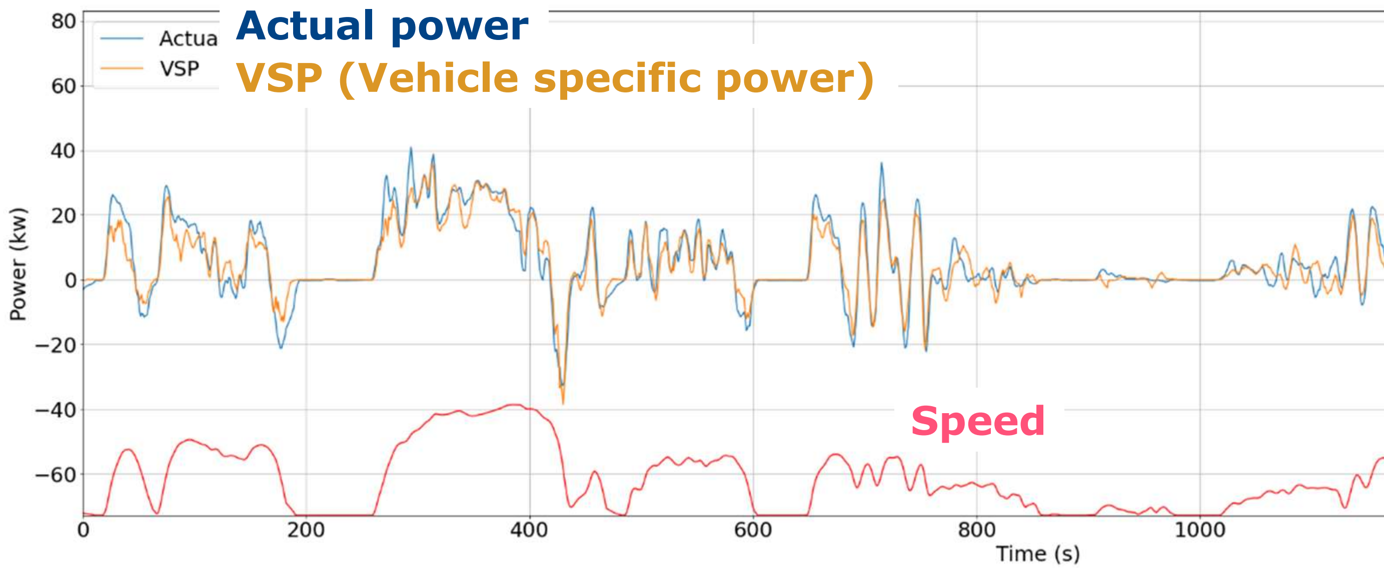
Data Set from E-Vehicles

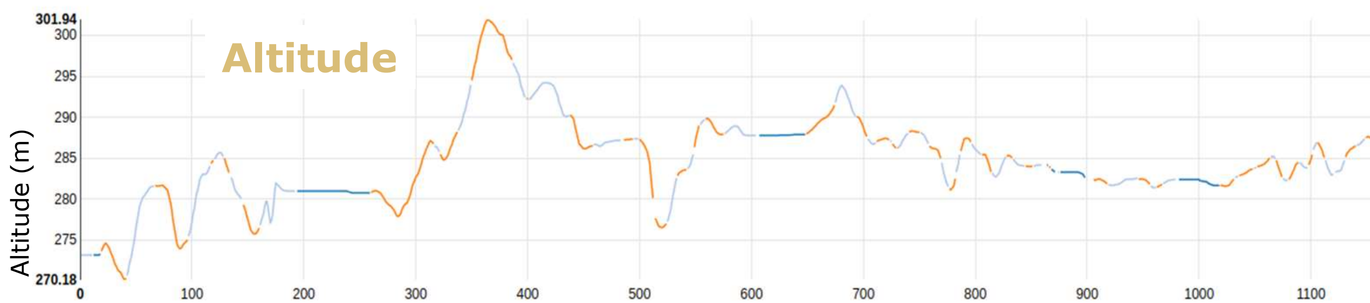
>500 trips, 8000km

Needed vs used Energy

- Calculate physically **energy needed** for movement, **„VSP"**
- Compare VSP to **actual power consumption**, for every second

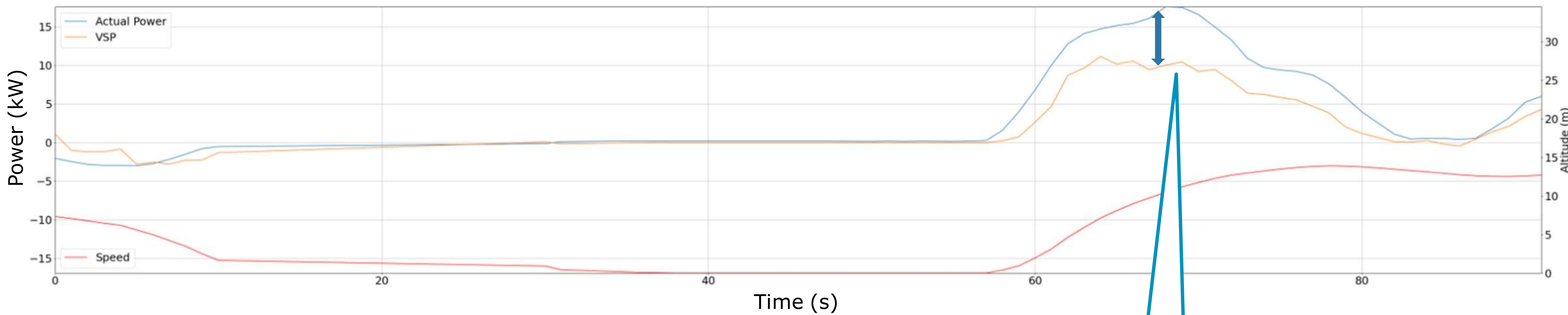
- Evaluation with Apache **Spark, batch processing**





Time (s)

Example in more detail: VSP vs Actual Power

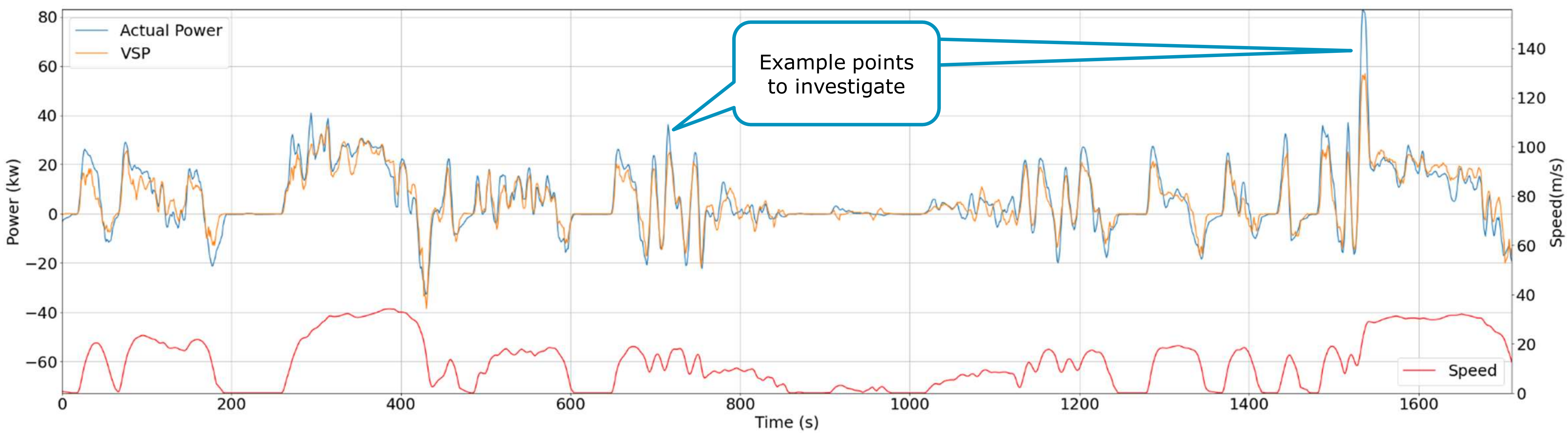


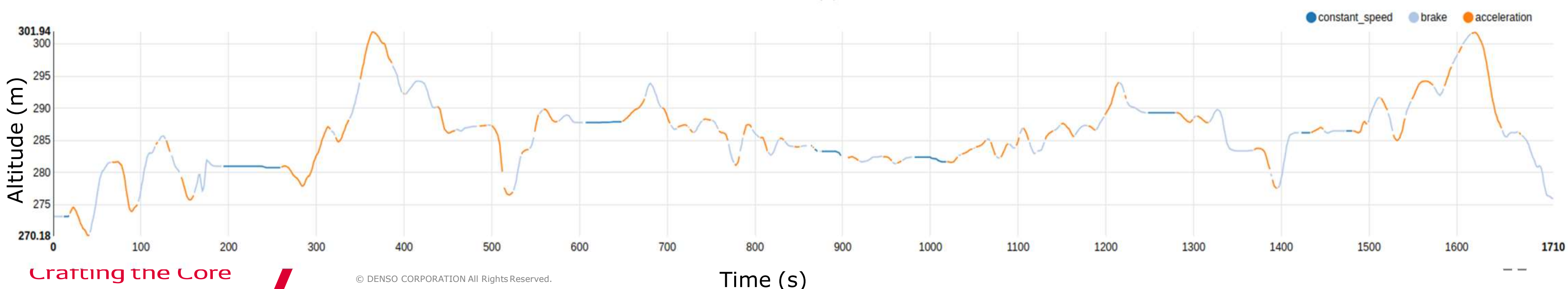
Gaps can indicate inefficiencies or issues

More Details:
Prehofer, C., & Mehmood, S. (2020, December). Big data architectures for vehicle data analysis. In *2020 IEEE International Conference on Big Data (Big Data)* (pp. 3404-3412). IEEE.

E-Vehicle Data with Uphill/Downhill

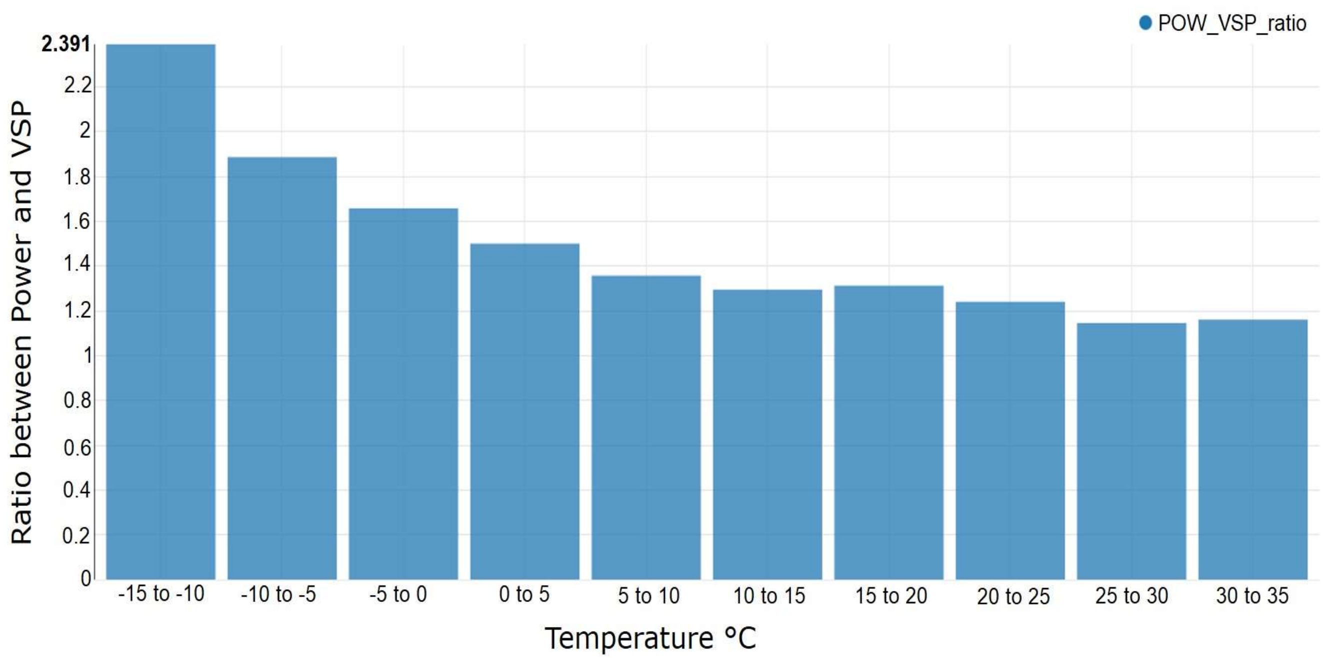
veh_trip_id	total_time	total_power_kW	total_VSP	Correlation(Power,VSP)	max_Temp °C	min_Temp °C	total_airConPow_kW	total_heaterPow_kW	POW:VSP
455_1720	1710.0	10476.059412500792	9865.72836081531	0.9481279556012671	15.0	13.5	0.0	0.0	1.0618637600148808





E-Vehicle Energy Consumption wrt Temperature

- Compute ratio between actual power and VSP for each trip
- Aggregation of 370 trips into temperature bins, total 4731 miles
- Clearly shows efficiency loss for colder temperatures



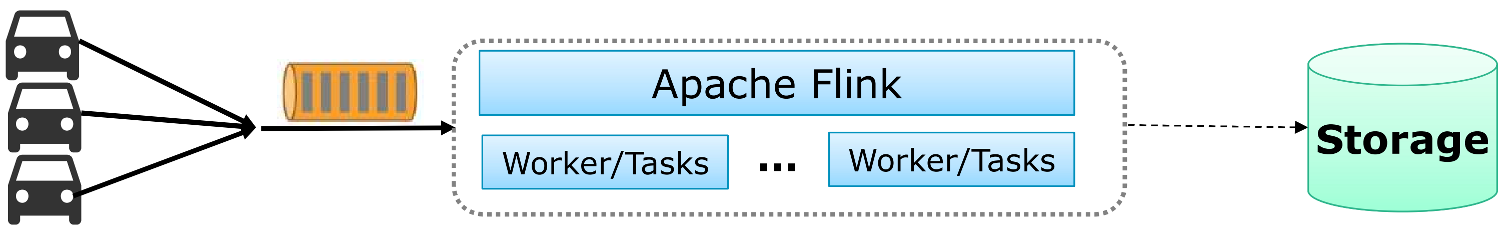
Recent work:
Using data to predict energy consumption for specific trips

Data Stream processing (for same use case)

Apache Flink Stream processing

- Apache Flink as a “true” streaming processing engine
- The core of Flink is streams and transformations on dataflows
  - Many APIs, incl DataStream and SQL

- Note: Apache Flink mainly designed for online stream processing, Spark for batch.



Performance: Flink Big Data Scales Down

		Vehicle data processing
Workstation, Intel Xeon W 3.7GHz, 8 Core, 3000 Euro	Number of Vehicle data streams (parallel)	**45k**
	Avgerage Latency range (ms)	1000 to 1800
Raspberry Pi 4b, ARM 7, 1.5 GHz, 4 Cores, 100 Euro	Number of Vehicle data streams	**12k**
	Avgerage Latency range (ms)	1000 to 2500

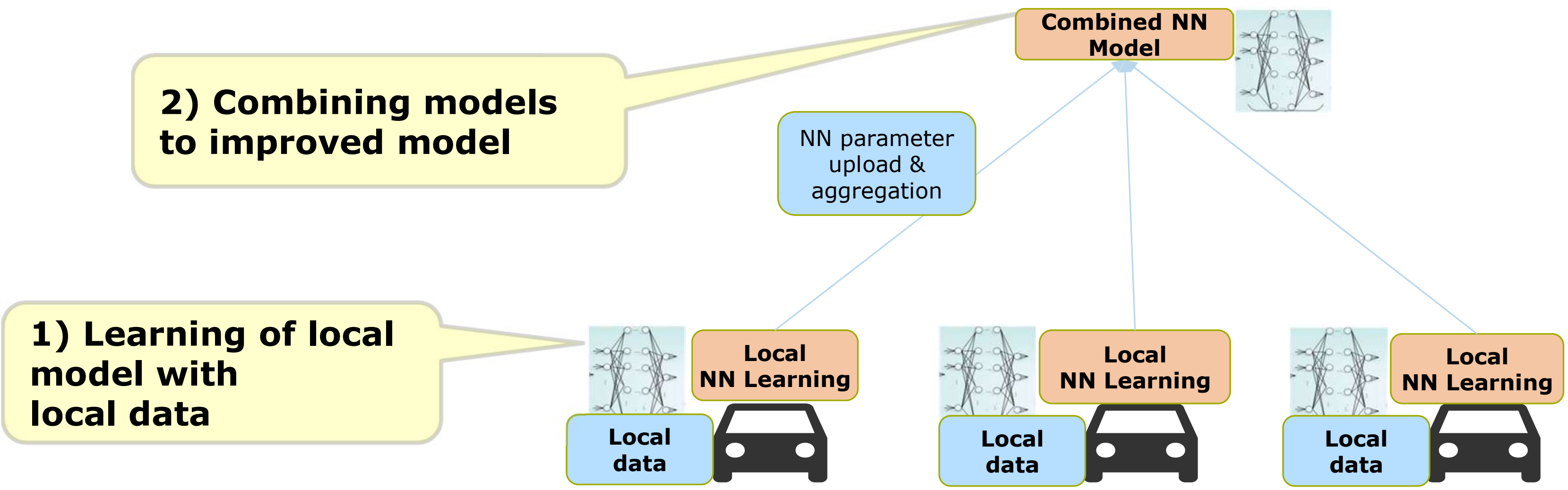
DENSO
Crafting the Core

Big Data for Vehicle Data Analysis

- **Big Data for connected vehicle applications**
  - Enormous amount of data
  - Many applications
- **Use case driving behavior & energy efficiency**
  - Compute efficiency for every second
  - **Comparison** of Big Data processing options
- **Use case driver status monitoring**
  - Privacy preserving data analysis with federated learning
- **Discussion and Outlook**

Federated Learning with Neuronal Networks (NN)

1. **Learning with local data** in cars to create local model
2. **Models are merged** from different vehicles/drivers (no image data upload!)
   - Exchange only NN parameters
3. **Improves privacy + data transfer volume**



Input Data: NTHU Dataset

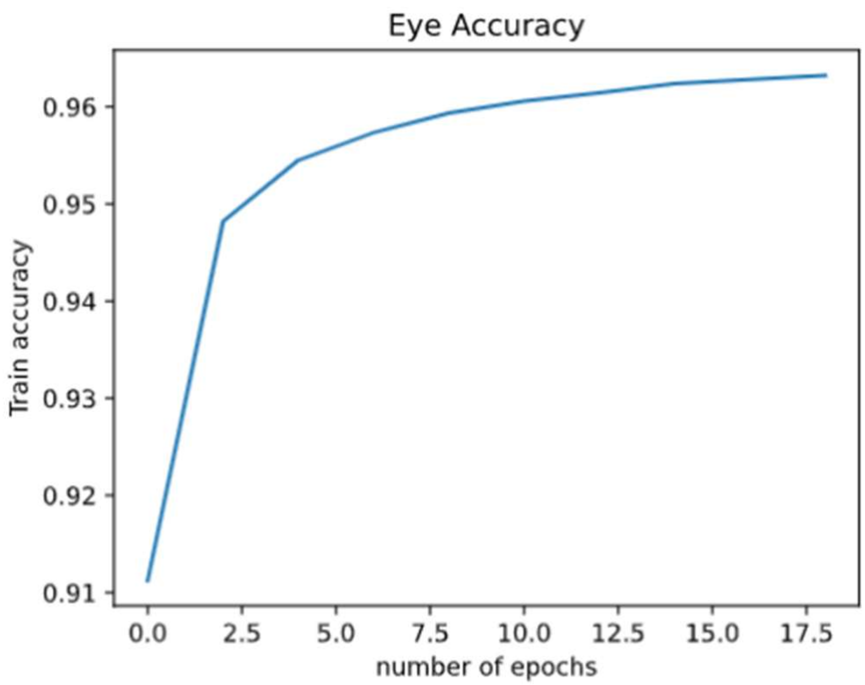
- 36 people of different genders and ethnicities
- Total 9 and a half hours
  (varying length videos)
- Annotated per frame
  (Eye, Mouth, Head, Drowsiness)
- Train, Val, Test Split (after preprocessing):

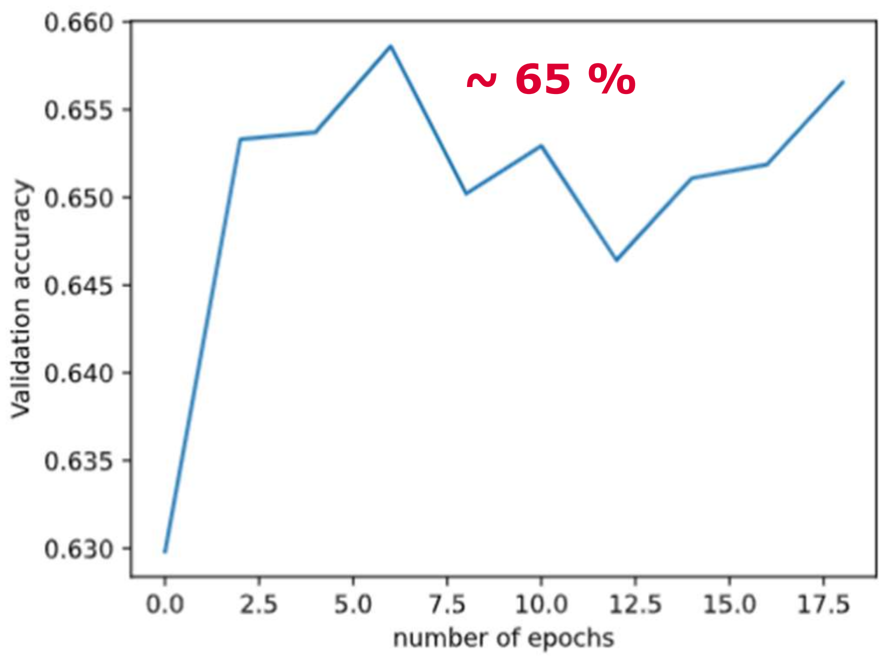
*	Training	Validation	Test
Number of Subject	18	4	14
Number of annotations (per-frame)	537,245	145,049	596,590
Number of Videos	288	16	56

Work done with TU München
Zafar, A., Prehofer, C., & Cheng, C. H. (2021, September). Federated Learning for Driver Status Monitoring. In *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)* (pp. 1463-1469). IEEE.

Dataset:
Weng et al., "Driver Drowsiness Detection via a Hierarchical Temporal Deep Belief Network."]

Baseline, Centralized Model, per frame accuracy

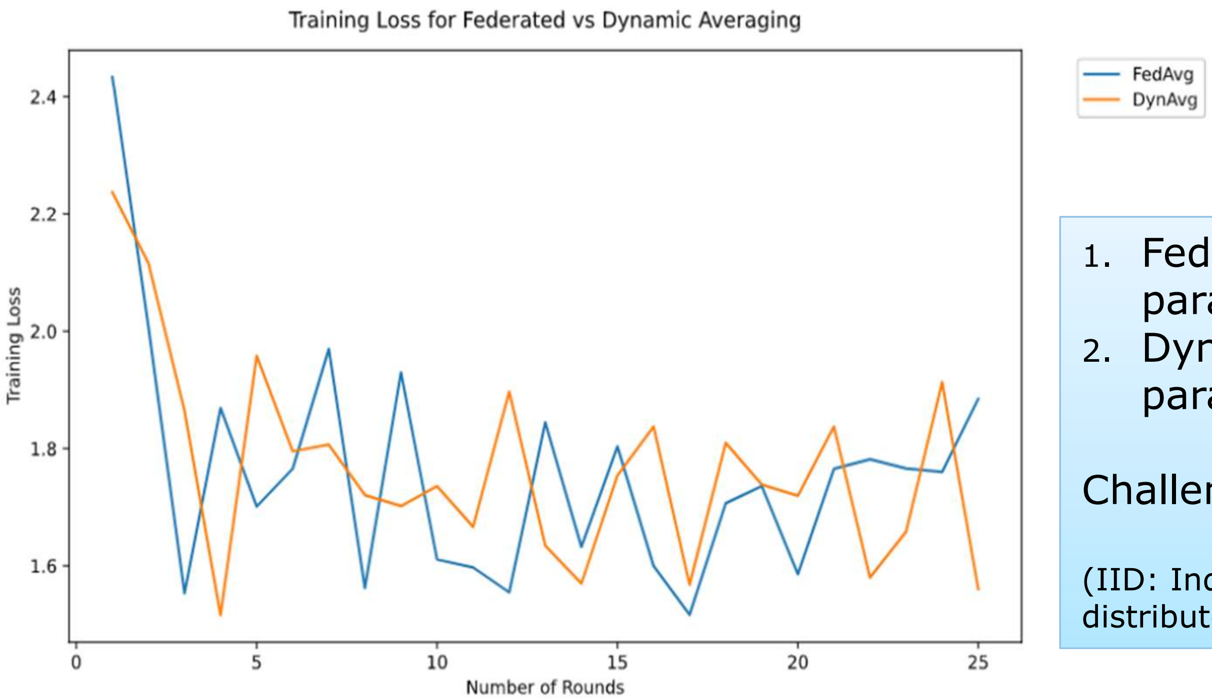




Note: Recognition of drowsiness based on **multiple frame results**

Experiment: initial learning rate: 1e −2 (0.01), lr decay: 0.001, momentum: 0.99, batchsize: 64, epochs: 20, batchNorm on conv layers + dropout rate (20%) on fc layers

Federated Learning: FedAvg and DynAvg



1. Federated Averaging: averaging of all parameters
2. Dynamic Averaging: significant parameter changes updated only

Challenges with highly non-IID data set!

(IID: Independent and identically distributed random variables)

Training loss from our experiment ($\Delta = 0.5$) shows no improvement with non-IID data

[Kamp et al., "Efficient Decentralized Deep Learning by Dynamic Model Averaging"]

Federated Learning: how much to aggregate from local updates

Promising Approach: FedProx
Regularization of loss function in local objectives to control divergence

Permit some local deviation from central model

Training loss from our experiment (μ = 0.01) shows improvement

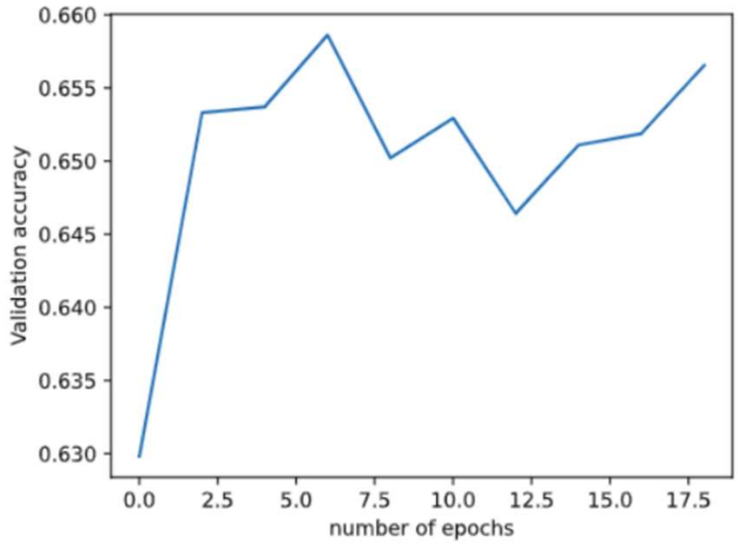
DENSO
Crafting the Core

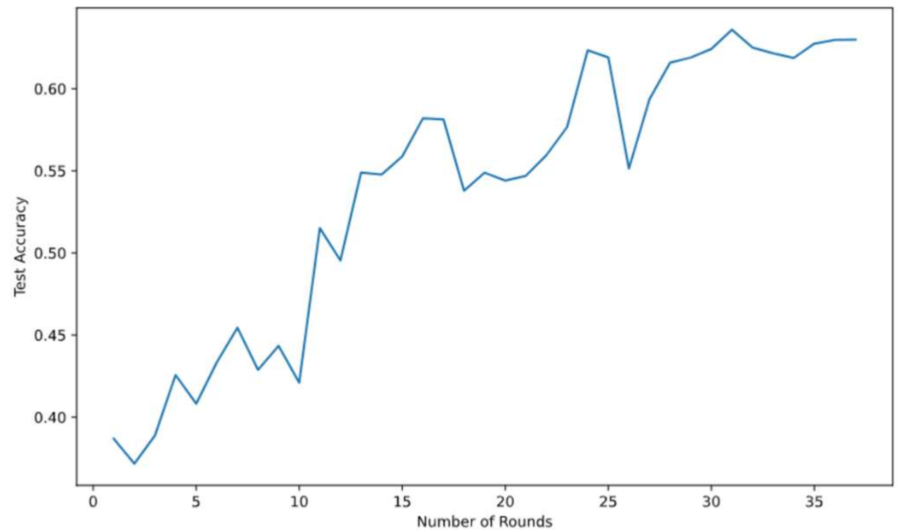
[Li et al., ""Federated Optimization in Heterogeneous Networks"]

Results: Comparing baseline vs FedProx

New results with >80% accuracy in our labs. Needed more and high-quality data

Predictive performance





Test Accuracy for baseline model (65%) and federated model (62%) for Eye Class

Big Data and Vehicle Data Analysis

Many applications for vehicle data
- **Different requirements**
- Challgenes of data **collection and processing**

Performance and scalability of Big Data solutions
- Apache Flink **scales down** to small machines (4 cores)
- **Distributed Big Data** processing can be highly efficient

Privacy-aware distributed AI with federated learning
- Training data does not leave the vehicle
- Promising first result on FedProx, currently ongoing work

Thanks to Shafqat Mehmood, Atiqa Zafar, Shumail Mohyuddin, Chih-Hong Chen, William Lindskog