IEEE ICC®

FedMFS: Federated Multimodal Fusion Learning with Selective Modality Communication

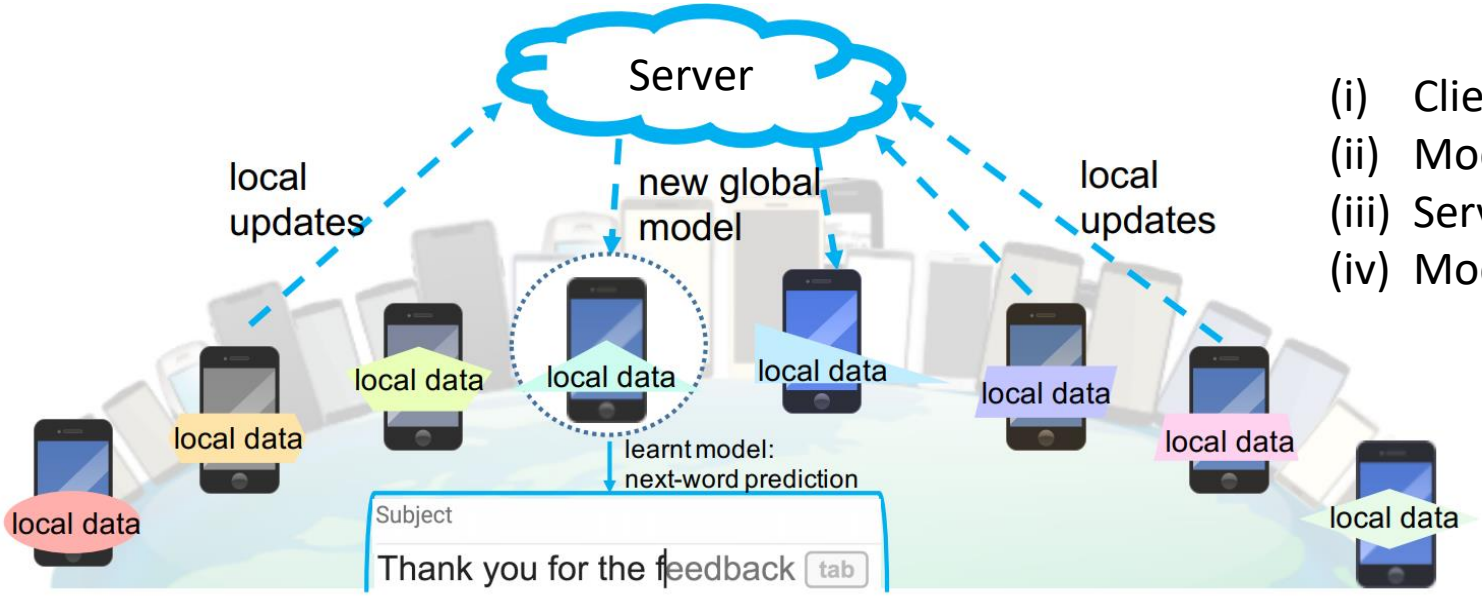
Liangqi Yuan, Dong-Jun Han, Vishnu Pandi Chellapandi,
Stanislaw H. Zak, and Christopher G. Brinton

June 10, 2024

PURDUE UNIVERSITY

Federated Learning

- **Goal:** Train either “a global model” or “personalized models” via collaboration



(i) Client-side model training
(ii) Model upload to the server
(iii) Server-side aggregation
(iv) Model broadcasting to clients

** Figure from SPM’20

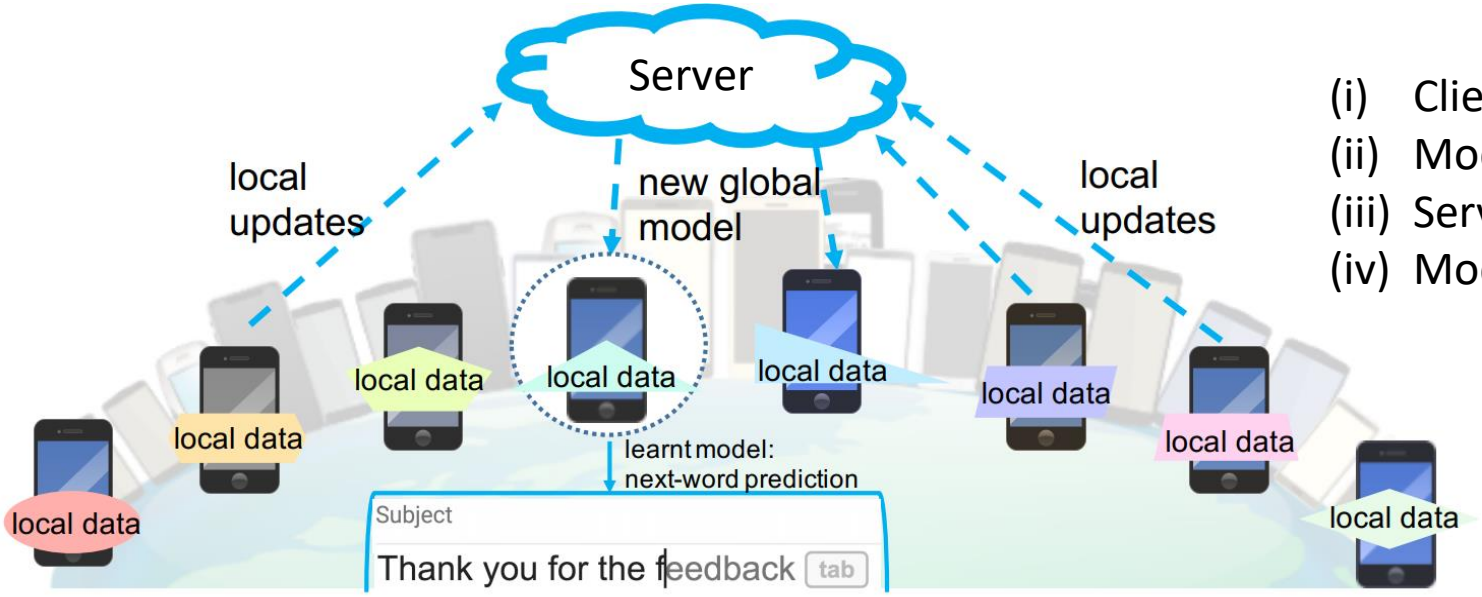
→ Communication burden issue, data heterogeneity issue, security and privacy issues etc.

→ Active area in both communications/networking and ML communities

[SPM’20] “Federated learning: Challenges, methods, and future directions,” IEEE Signal Processing Magazine, 2020.

Federated Learning

- **Goal:** Train either “a global model” or “personalized models” via collaboration



(i) Client-side model training
(ii) Model upload to the server
(iii) Server-side aggregation
(iv) Model broadcasting to clients

** Figure from SPM’20

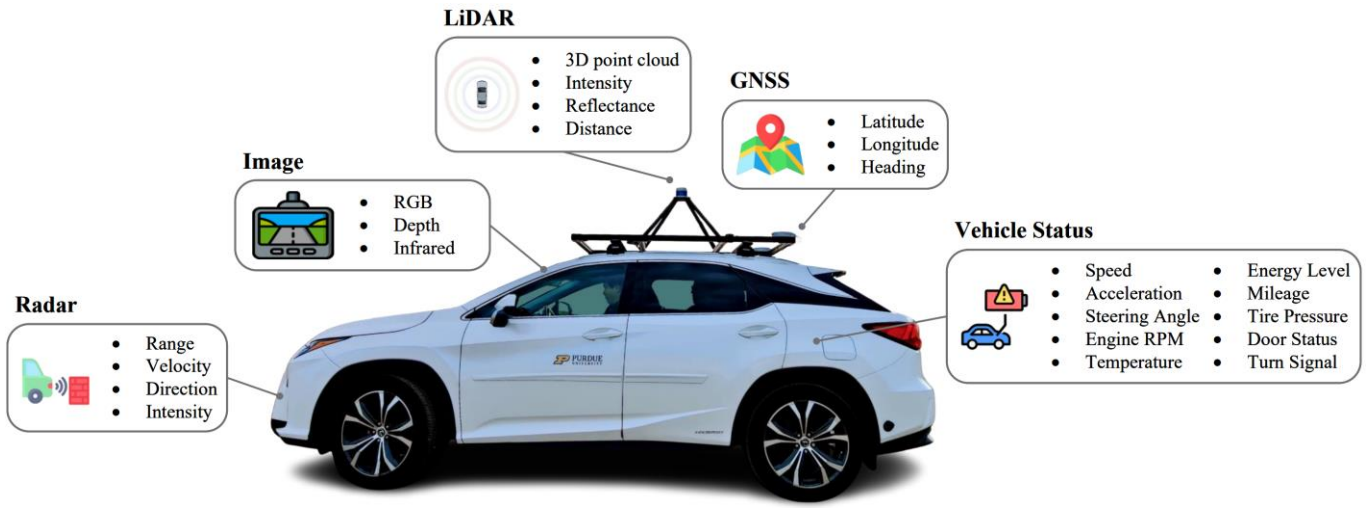
→ Communication burden issue, data heterogeneity issue, security and privacy issues etc.

→ Active area in both communications/networking and ML communities

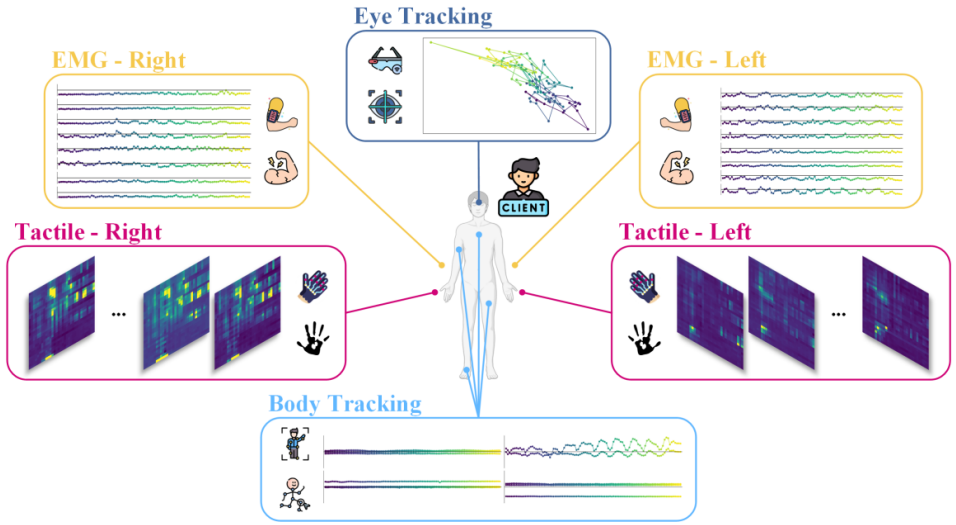
→ Most existing works consider a single-modality scenario (with a single type of data)

Motivation: Scenarios with Multiple Modalities

- In many practical settings, we need to make decisions based on multiple types of data
- Applications
  - **Autonomous vehicles:** Multiple types of sensors such as Cameras, LiDAR, and Radar
  - **Wearable sensors:** Multiple types of devices such as Smart Watches and Fitness Bands



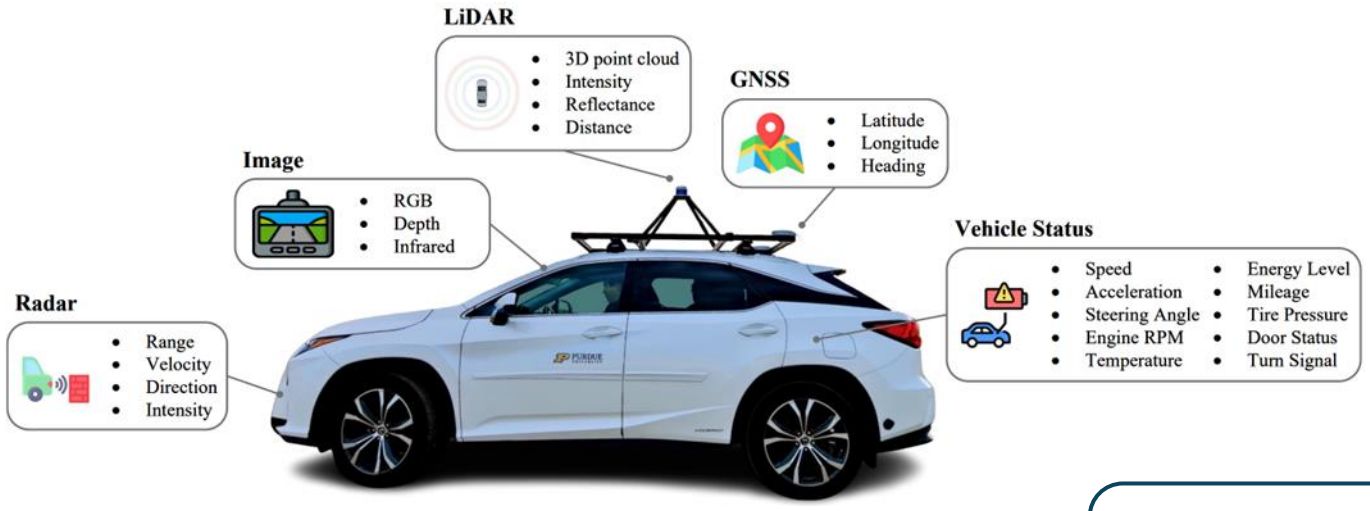
Autonomous vehicles



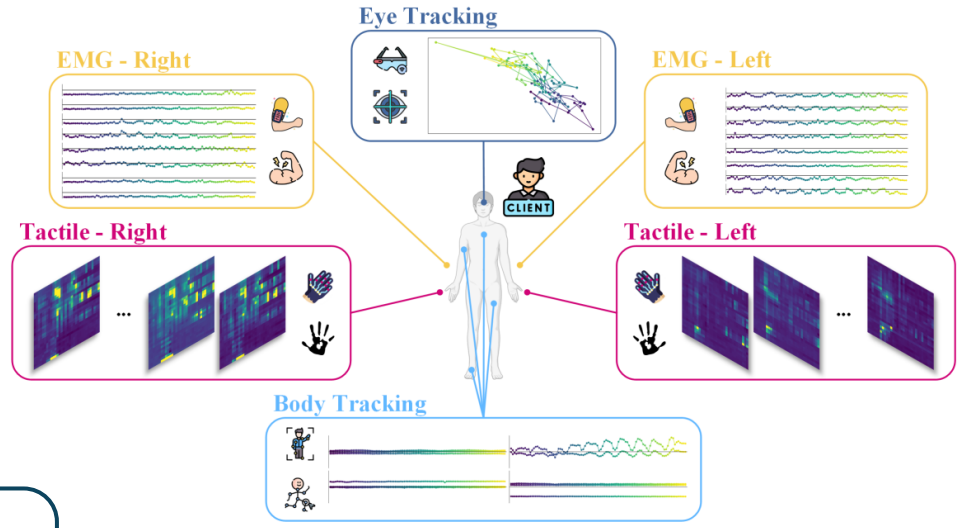
Wearable sensors

Motivation: Scenarios with Multiple Modalities

- In many practical settings, we need to make decisions based on multiple types of data
- Applications
  - **Autonomous vehicles:** Multiple types of sensors such as Cameras, LiDAR, and Radar
  - **Wearable sensors:** Multiple types of devices such as Smart Watches and Fitness Bands



Autonomous vehicles



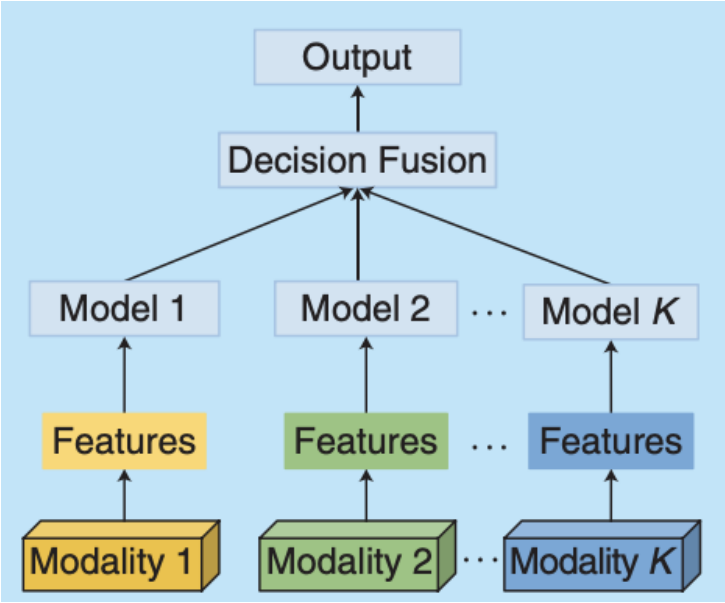
Wearable sensors

Multi-Modal Learning

Background: Multi-Modal Learning

- Multi-modal learning has been extensively studied in centralized settings
- One common approach is to use multiple models to process different types of data and combine them for decisioning

** Figure from SPM'17

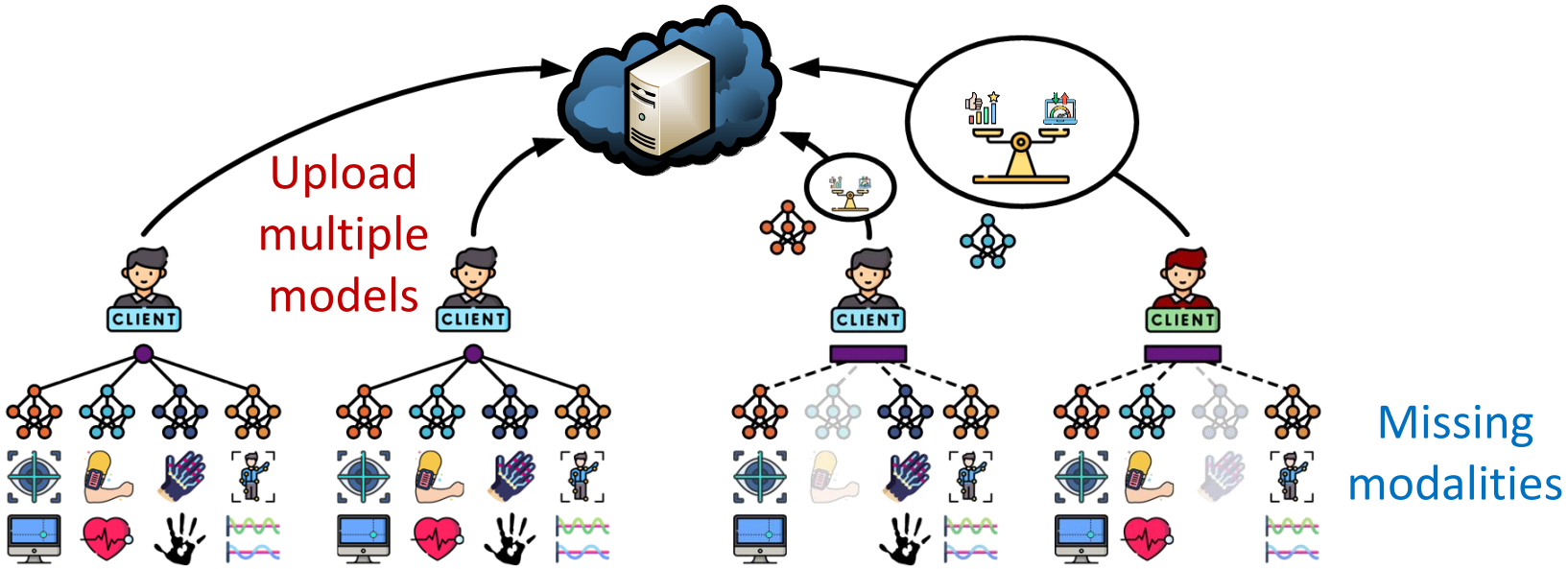


Multi-modal learning

[SPM'17] "Deep multimodal learning: A survey on recent advances and trends," IEEE Signal Processing Magazine, 2017.

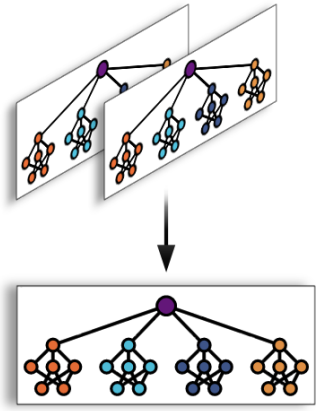
Our Focus: Multi-Modal Federated Learning

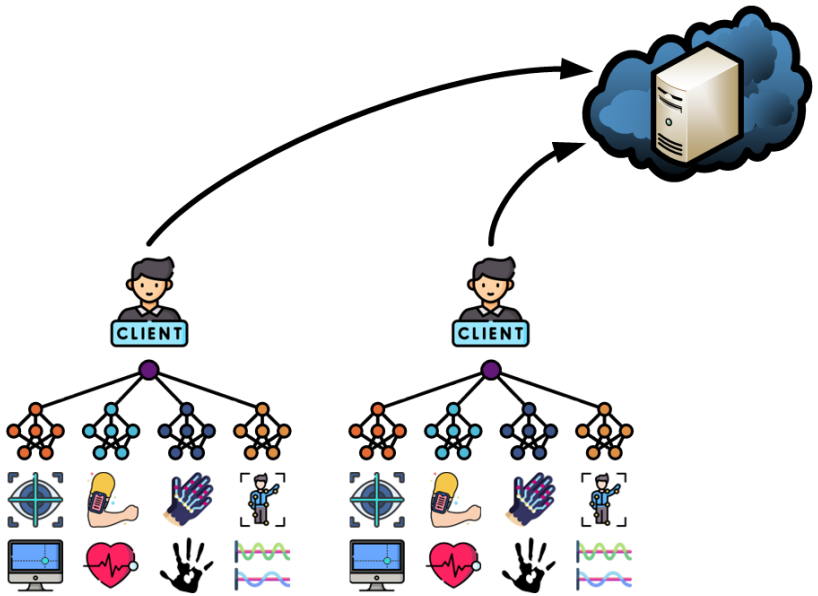
- Goal: Federated learning over distributed clients that have multiple modalities
- New challenges arise!
  - **Heterogeneous clients may lack certain modalities:** Some vehicles are not equipped with LiDAR, and some individuals with wearable sensors prefer not to wear watches
  - **Communication burden becomes more significant:** Each client needs to upload multiple models to the server



Multi-modal federated learning scenario

Overview



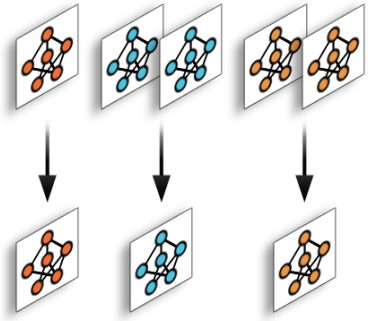


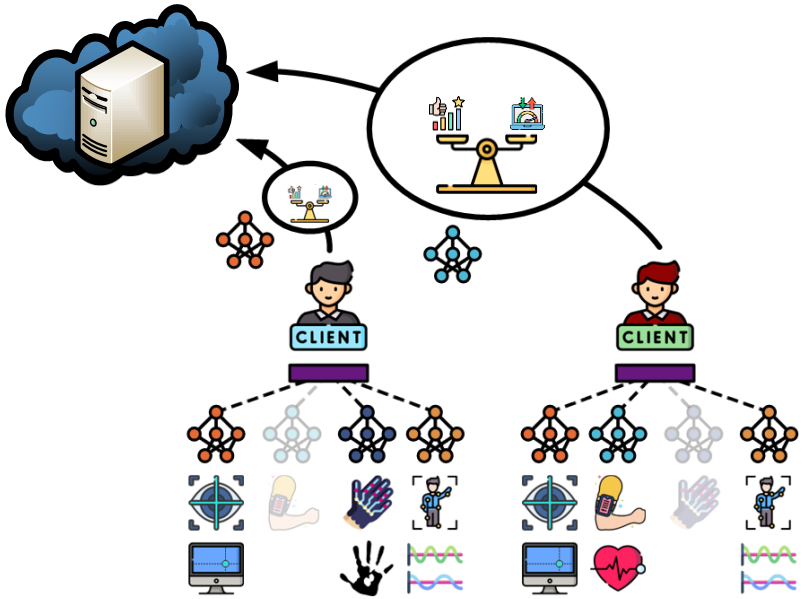
Existing Multimodal Federated Learning works:

Consider a scenario where all modalities are available

The entire model is uploaded to the server for aggregation.

Overview



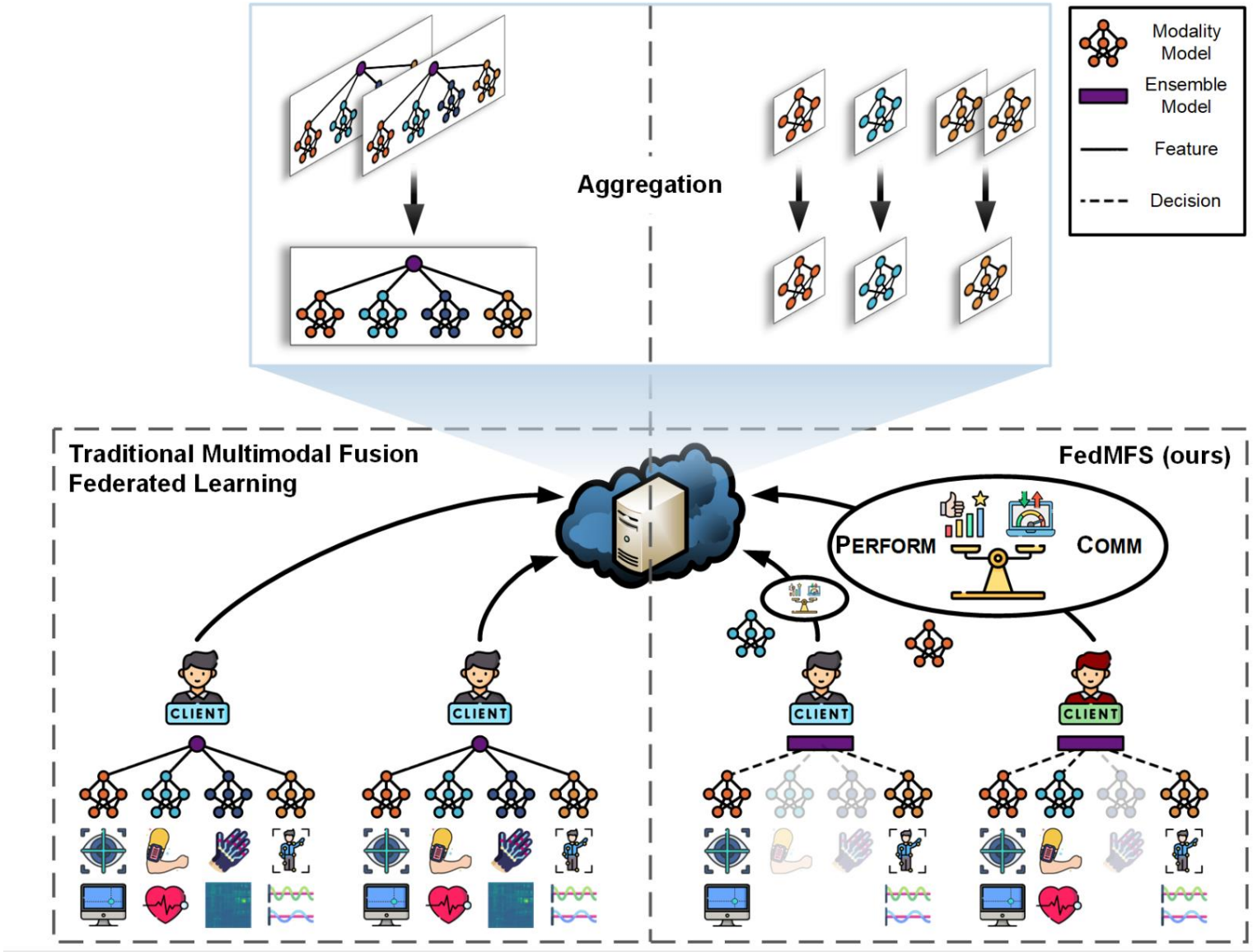


Our Approach:

Due to the decision-level fusion strategy, we allow clients to handle **missing modalities**

We significantly reduce the communication overhead through **selective modality communication**

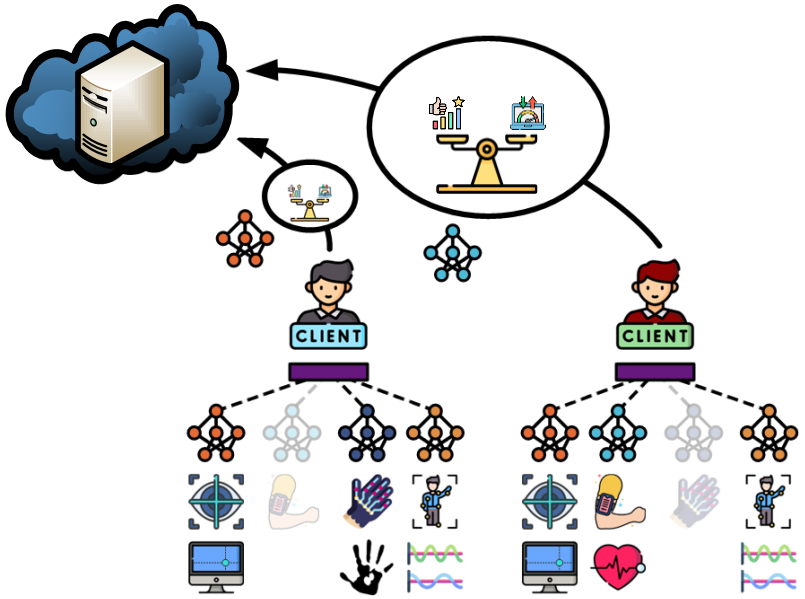
Overview



Selective modality communication

Selective Modality Communication

- FedMSF: **Fed**erated **M**ulti-Modal **F**usion Learning with **S**elective Modality Communication



Performance-Communication Trade-Off

- **Performance** – Shapley value (measures the impact of each modality)
- **Communication** – Modality model size

Priority measure of modality k = $\alpha_s \times \text{Shapley}_k + \alpha_c \times \text{Model size}_k$

- Weight of Shapley value α_s
- Weight of modality model size α_c

Each client **uploads** modality models based on **top-γ priority**

L

Priority

Xsens Awinda body tracking

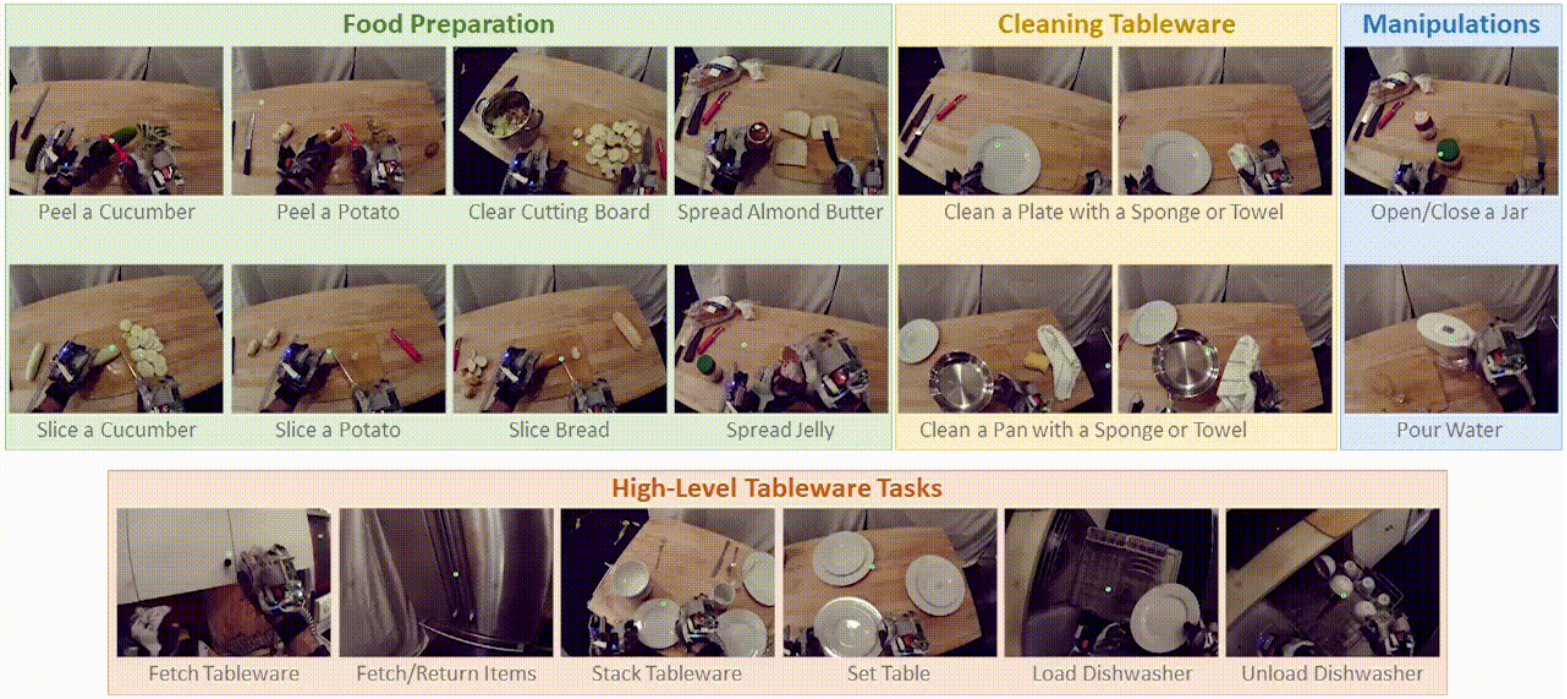
Manus Gloves finger tracking

Custom tactile sensors

Myo Armband muscle sensor

Pupil Labs eye tracking and first-person video

Classification of 20 classes in Kitchen activity



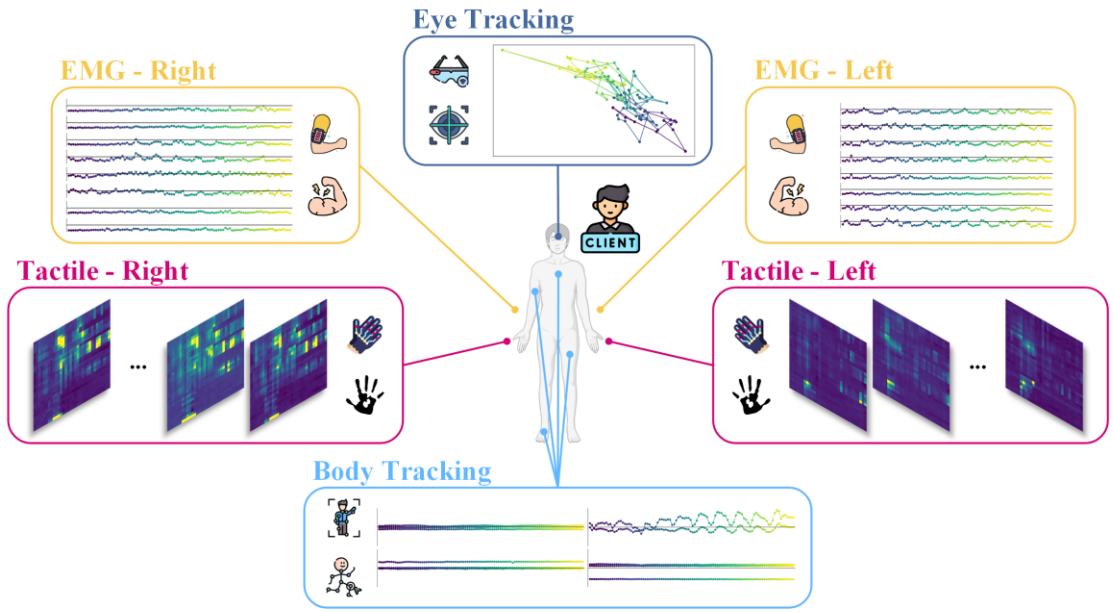


TABLE I
DESCRIPTION OF ACTIONSENSE DATASET

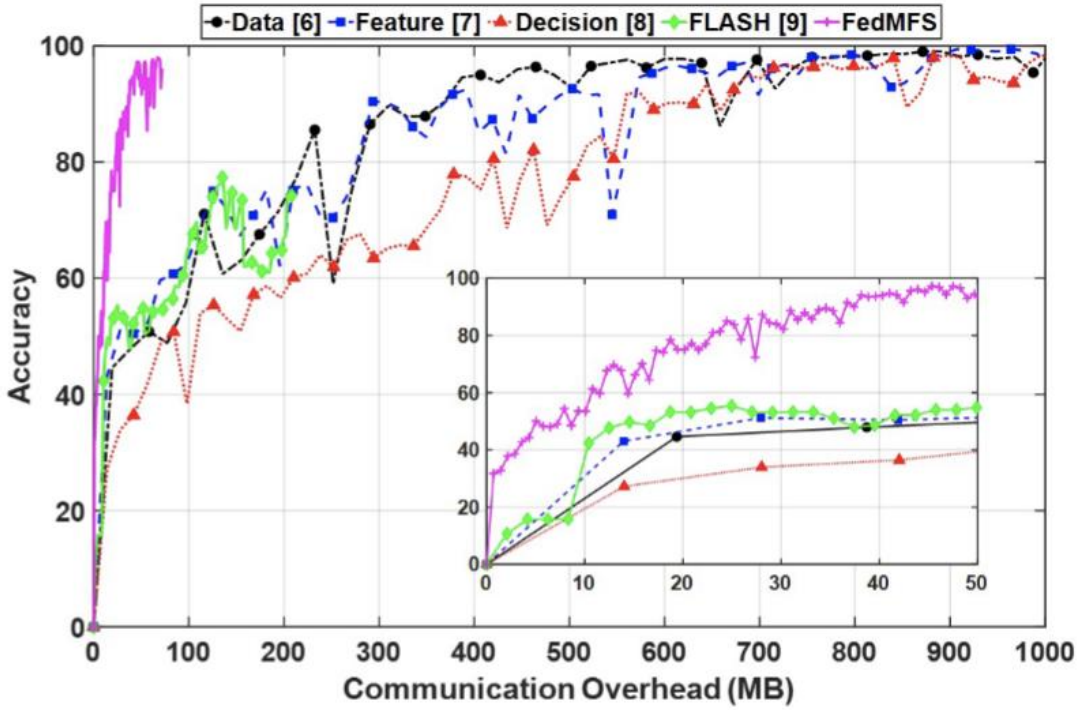
Sensor	Type	Position	Feature	Heterogeneity (Missing Data)
Eye Tracking	Position	Head	2	
Myo	EMG	Left Arm	8	
Myo	EMG	Right Arm	8	
Tactile Glove	Pressure	Left Hand	32×32	S06 – S09 [1]
Tactile Glove	Pressure	Right Hand	32×32	S06 – S09 [1]
Xsens	Rotation	Body	22×3	

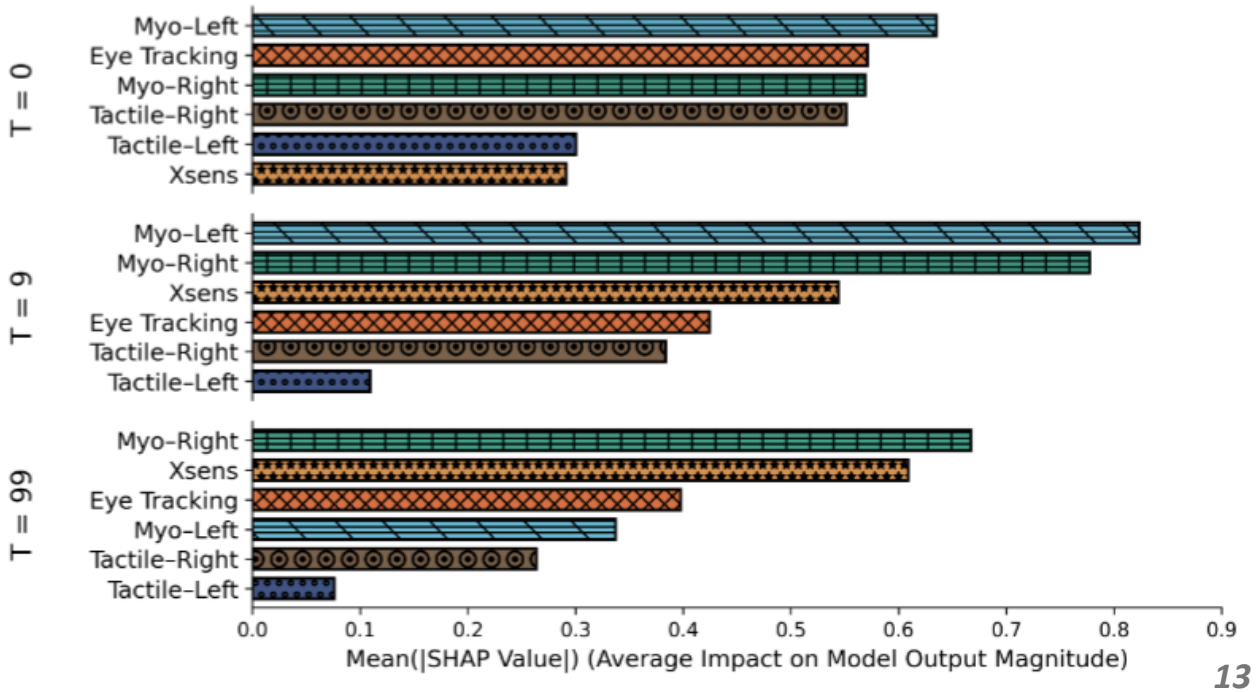
[1] S06 – S09 refers to subjects 06 through 09.

J. DelPreto, C. Liu, Y. Luo, M. Foshey, Y. Li, A. Torralba, W. Matusik, and D. Rus, "Actionsense: A multimodal dataset and recording framework for human activities using wearable sensors in a kitchen environment," *Advances in Neural Information Processing Systems*, vol. 35, pp. 13 800–13 813, 2022.

Experimental Results

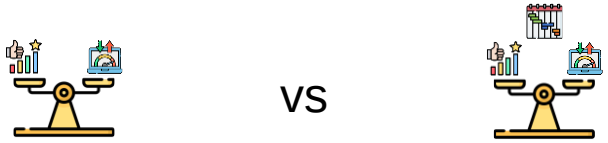
- Findings:
  - Our FedMFS reduces the communication overhead by over 4x.
  - Aggregation of all modality models is not always necessary.
  - Different data modalities contribute distinctively to the accuracy.



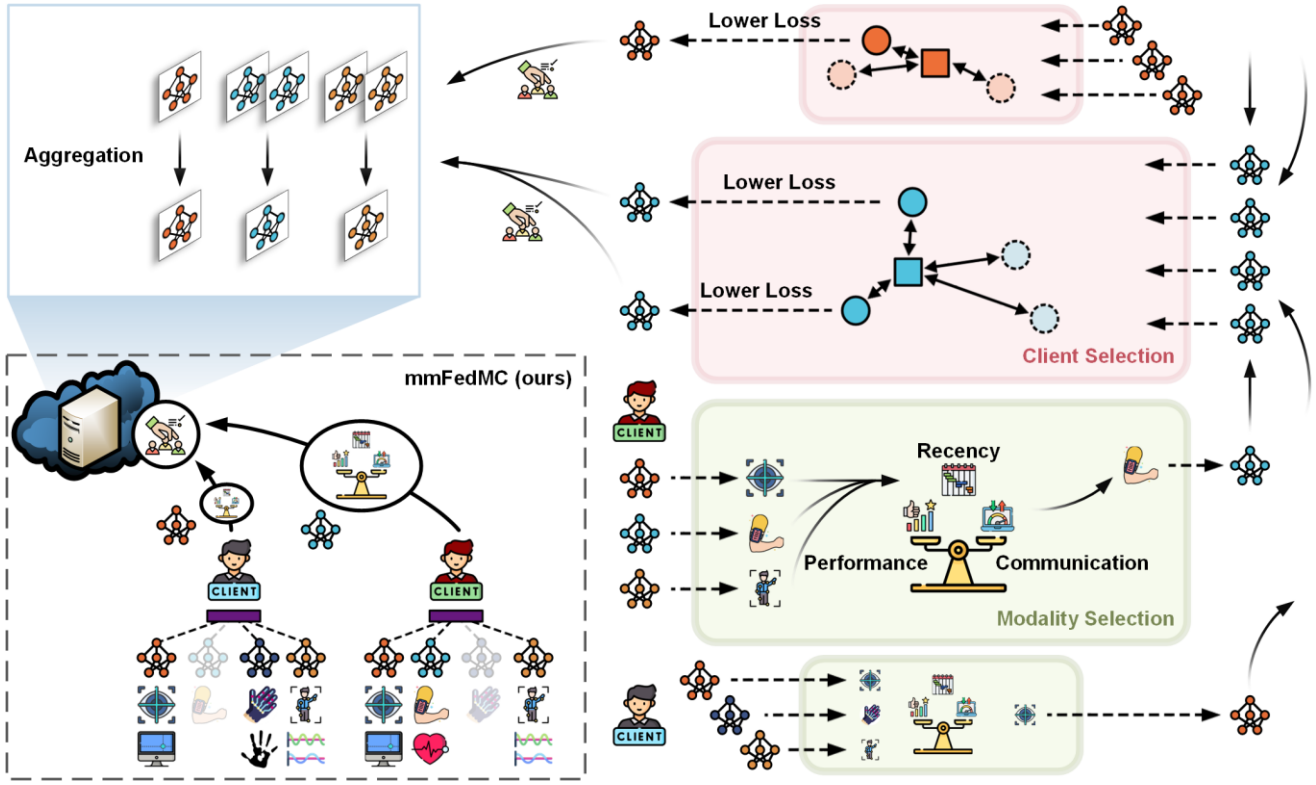


Extended Version

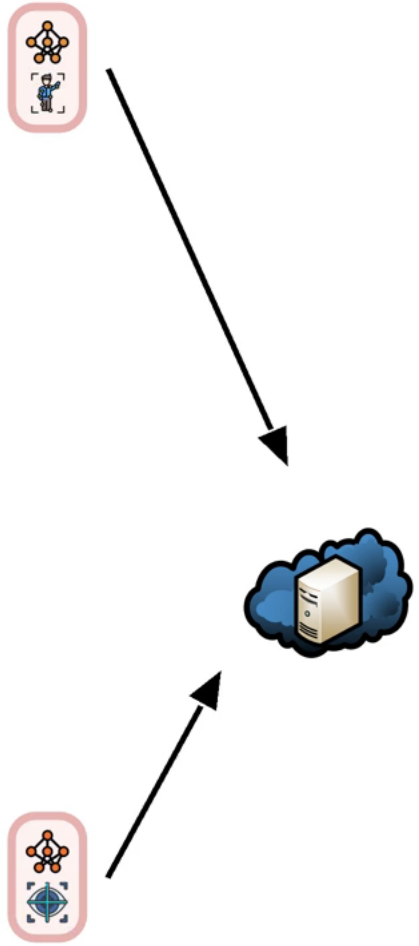
- **New metric**, *recency*, to prevent overemphasis on certain modalities and maintain generalization.

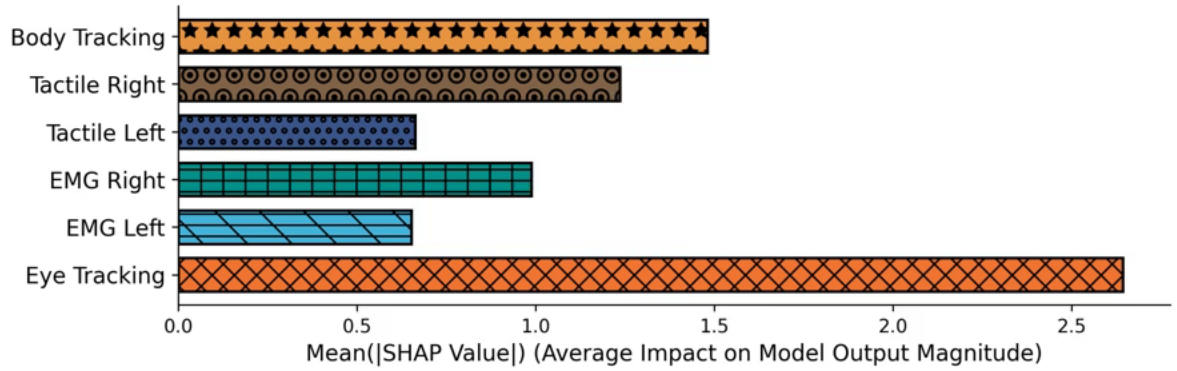


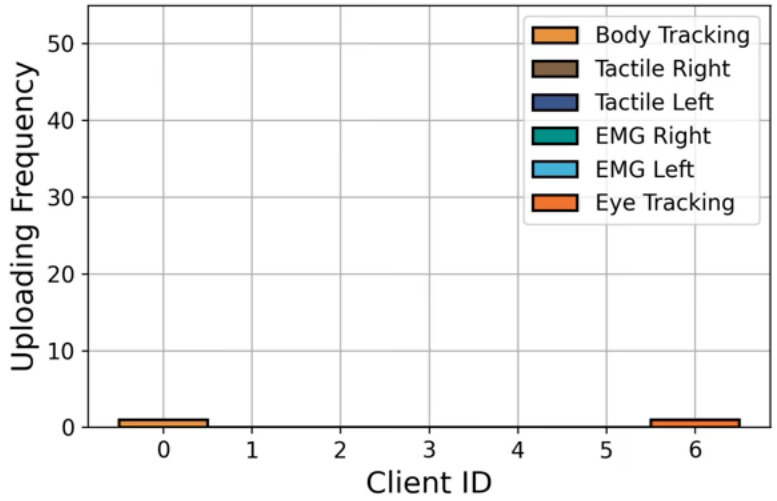
- **Client selection**, optimizing communication overhead in conjunction with modality selection.



Communication Round = 0

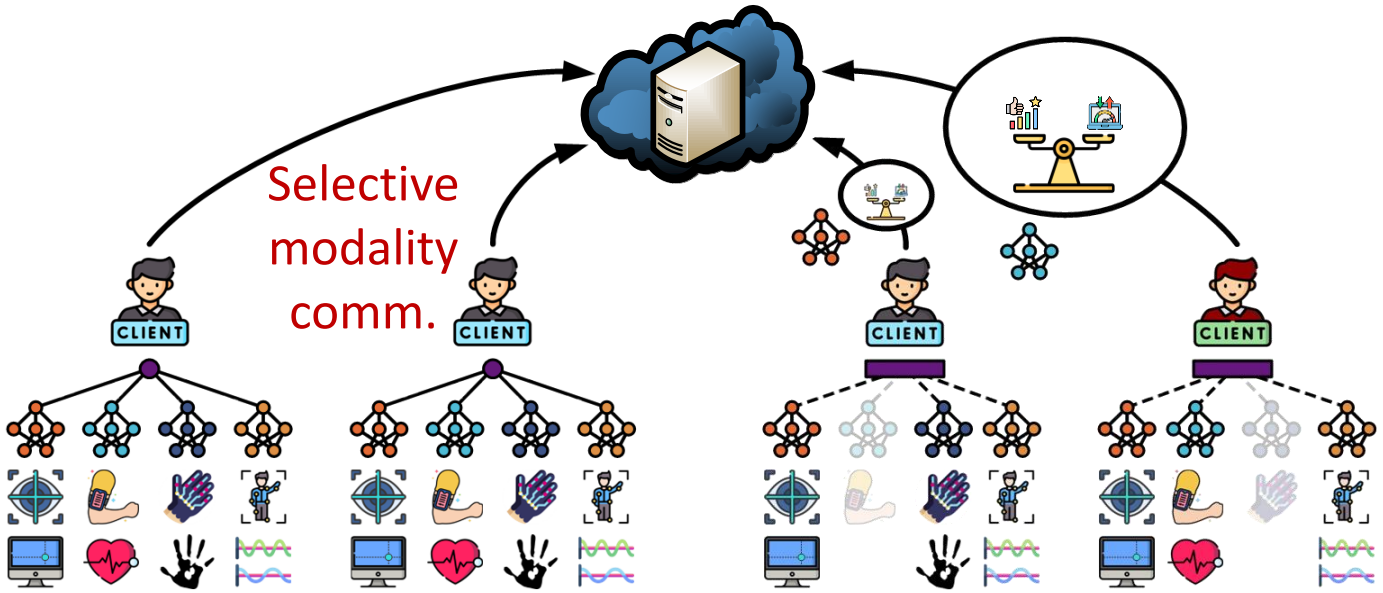






Conclusion

- Federated learning in a multi-modal scenario
- Consider missing modalities
- Selective modality communication to save communication burden while achieving a satisfactory accuracy



Thank you!

Liangqi Yuan - *liangqiy@purdue.edu*
Dong-Jun Han - *han762@purdue.edu*
Vishnu Pandi Chellapandi - *cvp@purdue.edu*
Stanislaw H. Żak - *zak@purdue.edu*
Christopher G. Brinton - *cgb@purdue.edu*

Conference Version

Extended Version
Demo Available!