Optimizing Cyber-Physical Systems with a Novel Automated Data Processing Pipeline

University of Kentucky

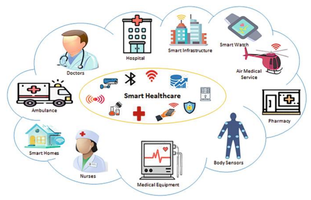
Presenter: Afsaneh Mahanipour

Advisor: Dr. Hana Khamfroush

Computer Science Department

Introduction and Motivation

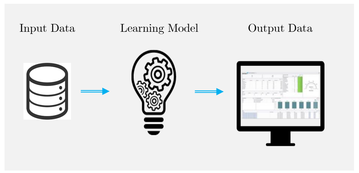
➢ Entering the era of Big Data in distributed environment



Smart healthcare system

Smart home system

Intelligent transportation system



Processing data overview

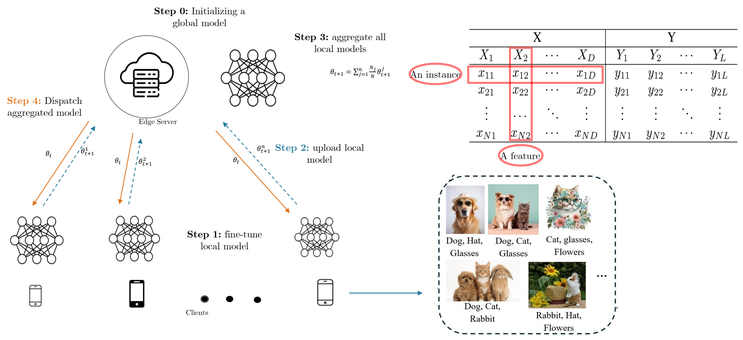
CHALLENGE

1. Presence of irrelevant, noisy, or redundant data.
2. Machine Learning and Deep Learning techniques are computationally expensive and training them is resource-intensive.
3. Resource constraints of IoT devices

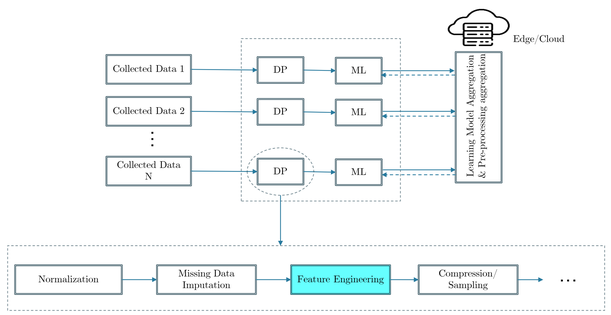
Using Data Pre-Processing techniques

- Select/construct informative features
- Reduce data size

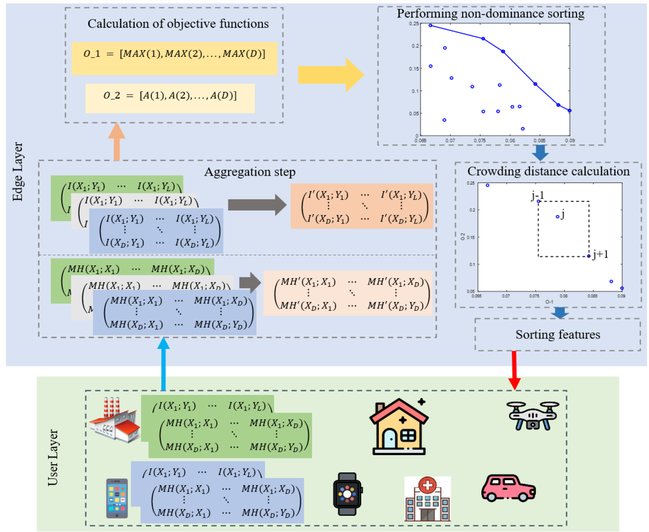
Methodology



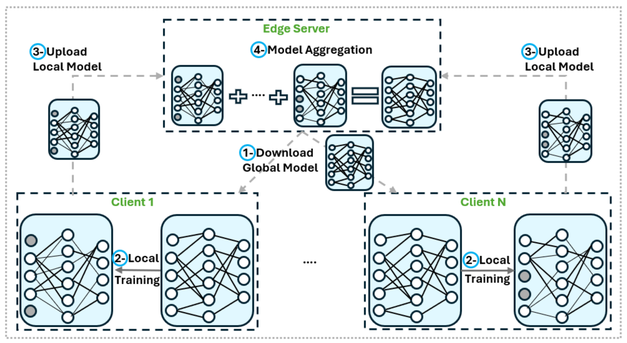
Architecture of Federated Learning

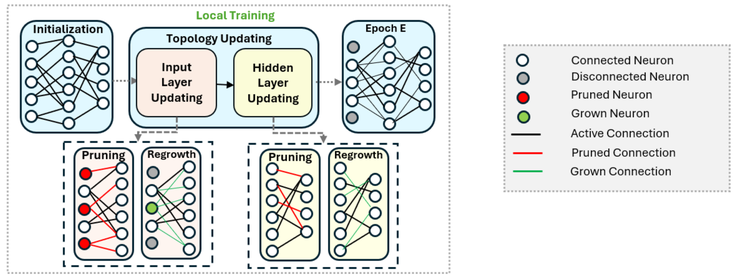


Integrated federated data pre-processing methods with machine learning models



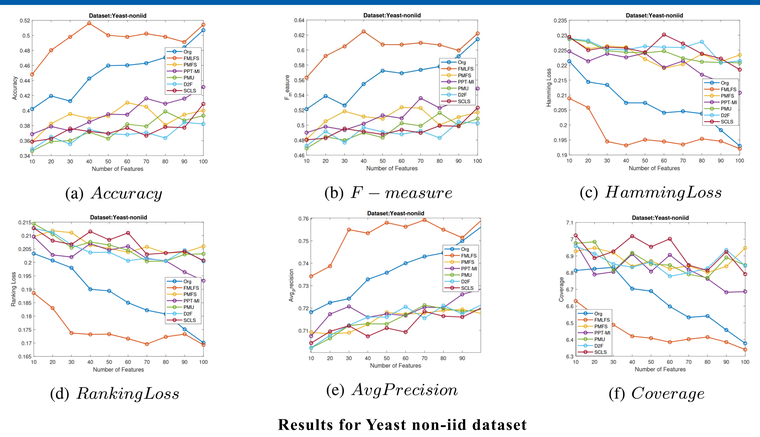
Overview of the FMLFS algorithm [2]





Overview of the proposed method DSFFS for embedded-based federated feature selection [7]

Results

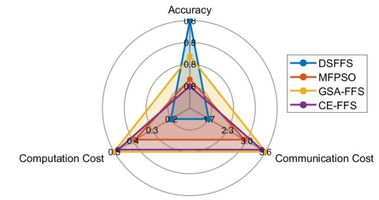


(a) *Accuracy* (b) $F-measure$ (c) *HammingLoss*

(d) *RankingLoss* (e) *AvgPrecision* (f) *Coverage*

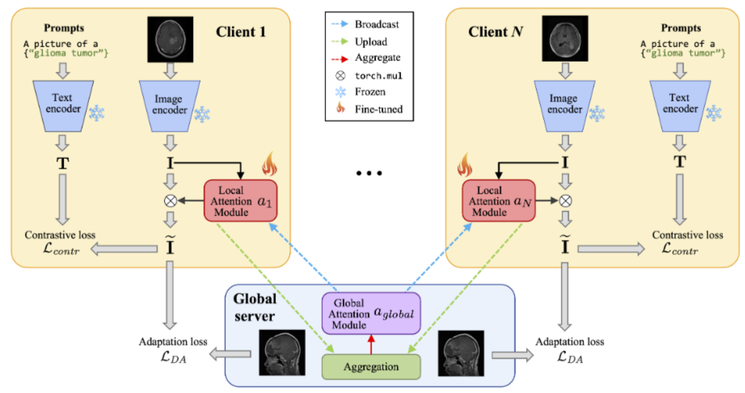
Results for Yeast non-iid dataset

➢ Integrating dynamic pruning and regrowing into the input layer of a sparse FL algorithm to perform feature selection.



(e) *MNIST*

Ongoing Works



Overview of the federated CLIP with feature attention module [8]

Conclusion

❖ In these works, wrapper-based, filter-based and embedded-based federated feature selection methods with different techniques are proposed to provide a good trade-off between learning model performance and communication-computation costs.

❖ The experimental results on different benchmark datasets from various domains demonstrate that the proposed methods beneficial for selecting informative features and providing suitable and small enough feature subsets.

❖ For example, in the Yeast dataset, the FMLFS method achieves an accuracy of 0.48 with only 20 features which is comparable to the performance of the classifier using 100 features without feature ranking on the cloud server, and other FS methods.

References

1. Mahanipour, A., & Khamfroush, H., “Fuzzy Federated Multi-Label Feature Selection: Reinforcement Learning and Ant Colony Optimization”, IEEE Big Data 2024, IEEE International Conference on Big Data, 2024.
2. Mahanipour, A., & Khamfroush, H., “FMLFS: A federated multi-label feature selection based on information theory in IoT environment”, SMARTCOMP 2024, IEEE International Conference on Smart Computing, 2024.
3. Mahanipour, A., & Khamfroush, H., “Enhancing IoT Security: A Novel Feature Engineering Approach for ML-Based Intrusion Detection Systems”, DCOSS-IoT 2024, 20th International Conference on Distributed Computing in Smart Systems and the Internet of Things, IEEE, 2024.
4. Mahanipour, A., & Khamfroush, H., “Multimodal Multiple Federated Feature Construction Method for IoT Environments”, GLOBECOM 2023, IEEE Global Communications Conference, 2023.
5. Mahanipour, A., & Khamfroush, H., “Wrapper-Based Federated Feature Selection for IoT Environments”, International Conference on Computing, Networking and Communications (ICNC2023), IEEE, 2023.
6. Mahanipour, A., & Khamfroush, H., “Semi-Supervised Federated Feature Selection for Multi-Label Data Using Fuzzy Information Theory”. (Under Review)
7. Mahanipour, A., & Khamfroush, H., “Embedded Federated Feature Selection with Dynamic Sparse Training: Balancing Accuracy-Cost Tradeoffs”. (Under Review)
8. Wu, Y., Desrosiers, C., & Chaddad, A. (2024, October). Facmic: Federated adaptative clip model for medical image classification. In International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 531-541). Cham: Springer Nature Switzerland

2025 Cyber-Physical Systems (CPS) Rising Stars Workshop, March 12, 2025, Vanderbilt University