Statistical Analysis for Applications of Federated Machine Learning using House Plants

Joshua M. Calzadillas
Advisors: Dr. MD Shaad Mahmud & Dr. John A. LaCourse
Department of Electrical & Computer Engineering, University of New Hampshire

University of New Hampshire

Introduction

Goal: To fundamentally understand and identify the benefits of a Federated Machine Learning (Fed-ML) system for a multisensory embedded system platform and to predict the state of a bamboo house plant using this model.

Motivation:

- No current research for houseplants is being used to study Fed-ML and the supporting framework.
- To design and operate a multi-sensor & multi-node system for a functional AI/ML predictor for house plants.
- Simple linear-based systems are limited in changing environments and require more functional resources to operate outside of predefined boundaries.

Background:

- Simple linear approaches are not sufficient for more complex scenarios.
- Preserving data privacy on the local device.
- Reducing unnecessary data to be aggregated into the model.

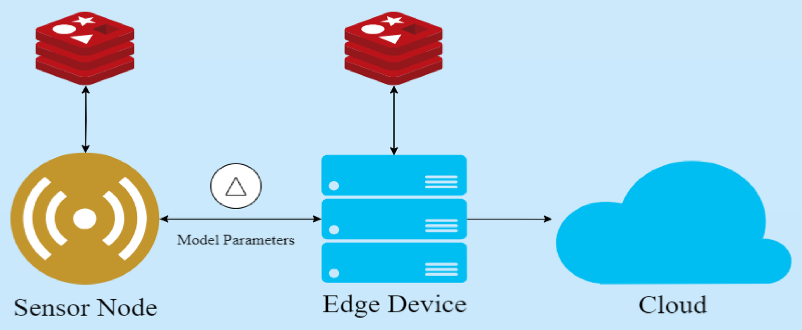
Novelty

- Approached a unique subject with a novel approach.
- Investigated a specialized area with the understanding of its applications towards AI & IoT.
- Focuses on the approach of minimalizing the data aggregation by localizing data.
- Reduces security risks by focusing on the AI model for general users.

Design Objectives

Infrastructure:

- High data integrity and reliability.
- Support for scalability (horizontal, vertical, and diagonal) for both sensor-nodes & edge-devices.
- Support for Horizontal Feature Space (due to the application).
- Incorporates a proper backup system for fault tolerance (checkpoint based).



Generalized architecture for interaction of sensor nodes to edge device for AI model training.

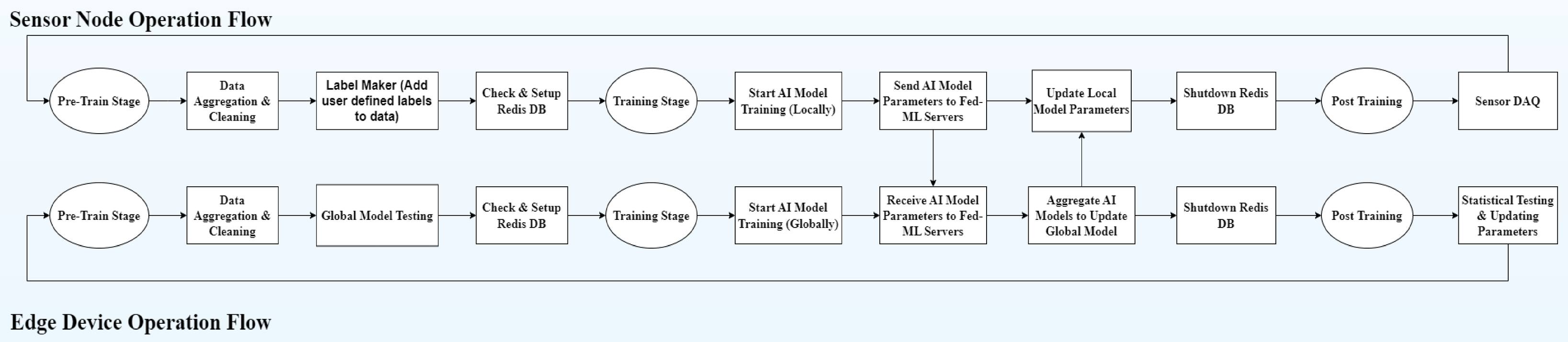
Sensor Node:

- Efficiently collecting multiple sensor datapoints in real time without losing data security.
- Have an integrated Fed-ML Ai system that incorporates features from a local environment.
- Build a predictor model that can account for the different states of chlorophyll levels & non-leaf items.

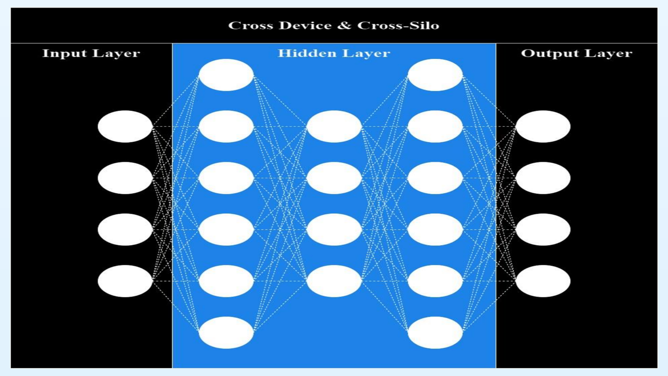
System Design

System Operation Flow:

1. Pre-Training Stage: Data cleansing, label definitions, and DB configuration.
2. Training Stage: Localized model training, Fed-ML model processing, updating model parameters.
3. Post Training: Sensor data collection (Sensor-Node) or Statistical Analysis of global model (Edge device).



System operation flow diagram (Sensor-Node for top, Edge device for bottom).



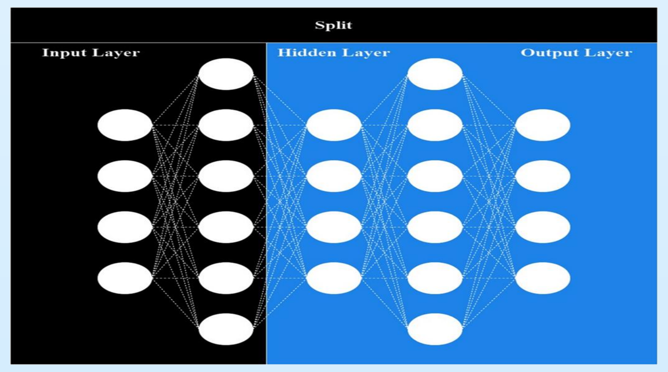
NN architecture focus for aggregation of Cross-Silo & Cross-Device (light blue background).

Cross-Device & Cross-Silo:

- The δ coefficients effects hidden-layer.
- Training process is straightforward because output information is localized.
- Training is done either on the edge device or the sensor-node.
- Requires more onboard memory and computational power for the sensor-node.

Split:

- The δ coefficients effects parts of the hidden-layer.
- Training process is more complex because output information is not localized.
- Training is done on both edge & sensor-node devices.
- Requires less onboard memory and computational power for the sensor-node.



NN architecture focus for aggregation of Split (light blue background).

Sensor-Node collecting data on plant using light sensors.

The sensor nodes have the capability to collect sensor data in real time using the python programming language and Linux based OS. Python was chosen because of it's vast support for AI, Data Science, and Raspberry Pi hardware interfacing libraries.

Sensor Node Resources:

- Raspberry Pi 4 & Zero W
- 3 or 4 ADPS9960 IR light sensors
- Temperature sensor
- I2C MUX
- 32 GB SD Card
- Raspbian OS

Plant Health Model & Results

The color difference between each leaf.

AI Predictor System:

- 3 primary states (H, M, L) correlate to the chlorophyll levels.
- Added non-leaf like features to account for situations where the light sensor does not collect leaf data.
- AI predictor mostly predicts between 2 states (H & M or M & L).

Model Accuracy:

When comparing the results of the different Fed-ML systems with each approach, the model accuracy is the most efficient way of evaluating a Fed-ML system.

Name	Theoretical Accuracy (Based on Papers)	Measured Accuracy
Baseline	X	81%
Cross-Device	~78% (average test) [2]	~80%
Cross-Silo	~ 50% (test) [3]	~80%
Split	~ 75% (validation) [4]	X

Fed-ML stats comparison with both theoretical & measured accuracy.

Conclusion

The importance of using a Fed-ML approach rather than a general AI or simplistic linear approach to this problem allows for a degree of freedom that can be used to improve the quality of this project by adjusting the number of sensors, neurons, nodes, and complexity of this project. The main goal of the sensor node is to predict the state of the plant based on features that the sensors can recognize, which in this case is the Chlorophyll level. The use of Fed-ML improves the capabilities that an application can utilize by expanding on dimensionality of the problem.

Future Work & Acknowledgements

- Implement GAN AI Architecture.
- Increase areas of focus for project.
- Apply to other plants.

- UNH ECE Department
- Dr. Shaad Mahmud
- Dr. John LaCourse

References

[1] N. Özreçberoğlu and İ. Kahramanoğlu, "Mathematical models for the estimation of leaf chlorophyll content based on RGB colours of contact imaging with smartphones: A pomegranate example," *Folia Horticulturae*, vol. 32, no. 1, pp. 57–67, Jun. 2020, doi: 10.2478/fhort-2020-0006.
[2] D. Chen *et al.*, "FS-Real: Towards Real-World Cross-Device Federated Learning." arXiv, Mar. 23, 2023. doi: 10.48550/arXiv.2303.13363.
[3] A. Das, T. Castiglia, S. Wang, and S. Patterson, "Cross-Silo Federated Learning for Multi-Tier Networks with Vertical and Horizontal Data Partitioning," *ACM Trans. Intell. Syst. Technol.*, vol. 13, no. 6, p. 99:1-99:27, Sep. 2022, doi: 10.1145/3543433.
[4] P. Vepakomma, O. Gupta, T. Swedish, and R. Raskar, "Split learning for health: Distributed deep learning without sharing raw patient data." arXiv, Dec. 03, 2018. doi: 10.48550/arXiv.1812.00564.