Preservation of Patient Confidentiality using Edge Computing and Distributed Machine Learning: An early Proof of Concept Study

Hugo Hiden, Paul Watson, Jamie McQuire, Nick Wright, Michael Catt

UK Systems Research Workshop 2021

Motivation

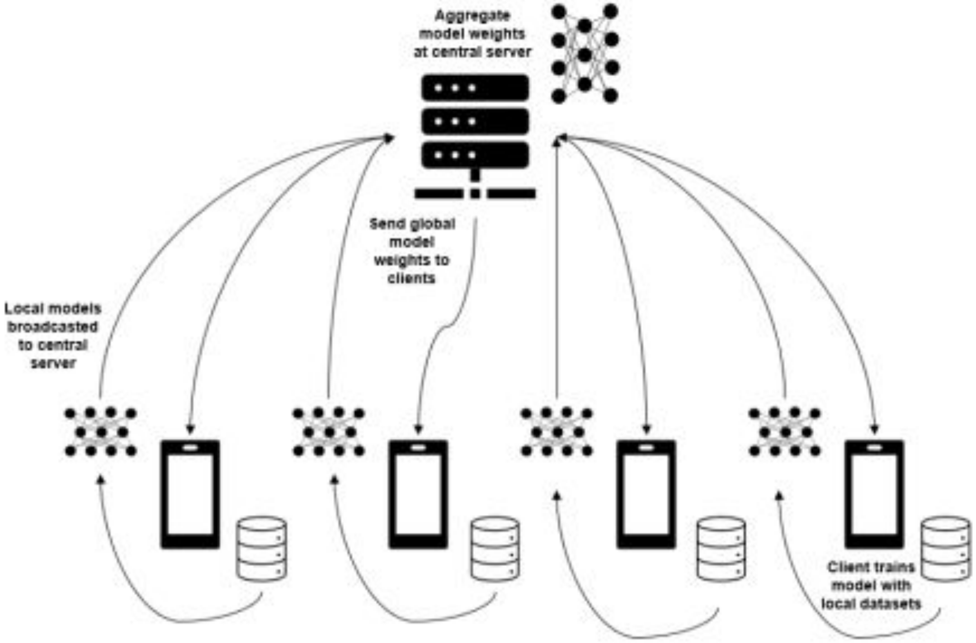
- IoT and Mobile technologies are generating massive amounts of data.
- GDPR and HIPAA regulations.
- Confidentiality and privacy of patients is important.
- Limited bandwidth of mobile devices can prohibit mass data uploads to cloud services.
- Improvements in Edge Computing has enabled ML algorithms to run locally i.e. TinyML and TensorFlow Lite.

Federated Learning

Distributed machine learning paradigm that involves a loose federation of clients jointly training a machine learning algorithm under the coordination of a central server.

- Privacy-preserving.
- Communication efficient.
- Uses decentralized data-sources.
- Cross-silo and cross-device.

Federated Averaging (FedAvg)


Aggregate model weights at central server
Send global model weights to clients
Local models broadcasted to central server
Client trains model with local datasets

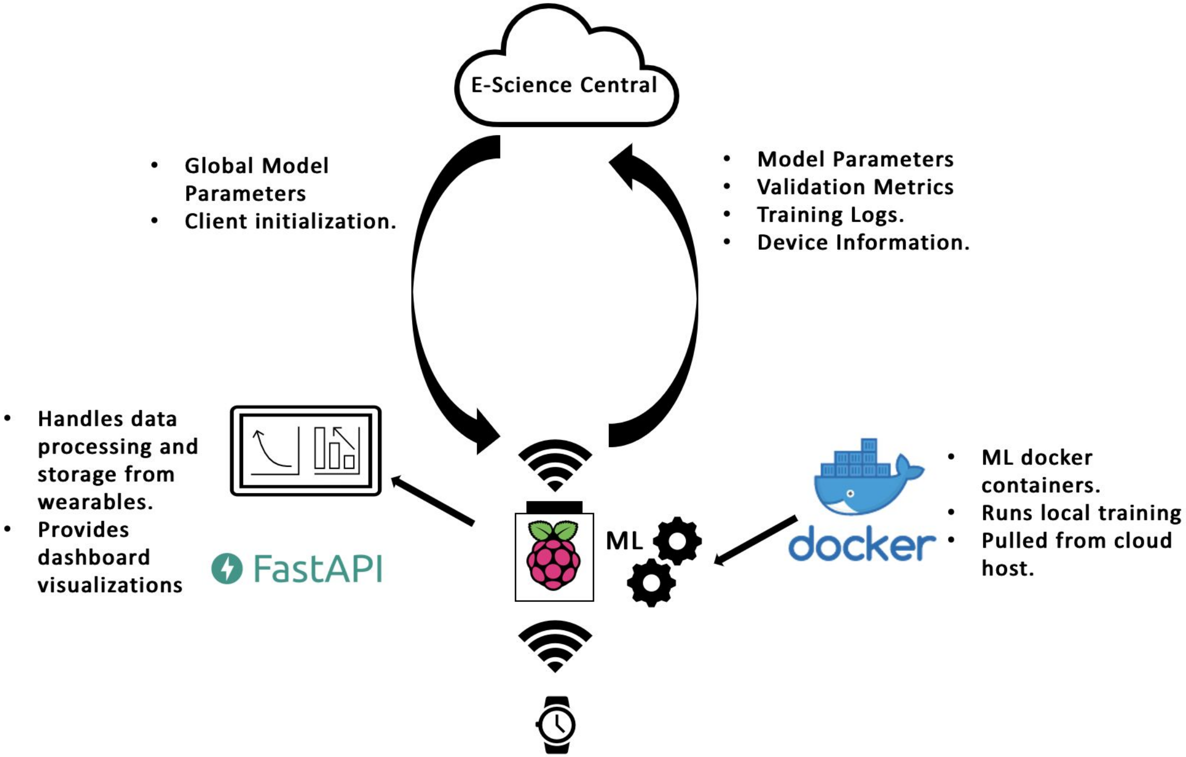
Federated Learning Systems

- Limited research in designing a system at-scale.
- Federated Learning software is primarily for simulations.
- Basic device - cloud implementations.
- Scalability and device management is a challenge.

Proposed System

- Multi-layer, Edge-Cloud architecture.
- 3 Sections:
    - Edge computing nodes.
    - e-Science central.
    - Cloud federated learning model.

Edge Computing



- Global Model Parameters
- Client initialization.

- Model Parameters
- Validation Metrics
- Training Logs.
- Device Information.

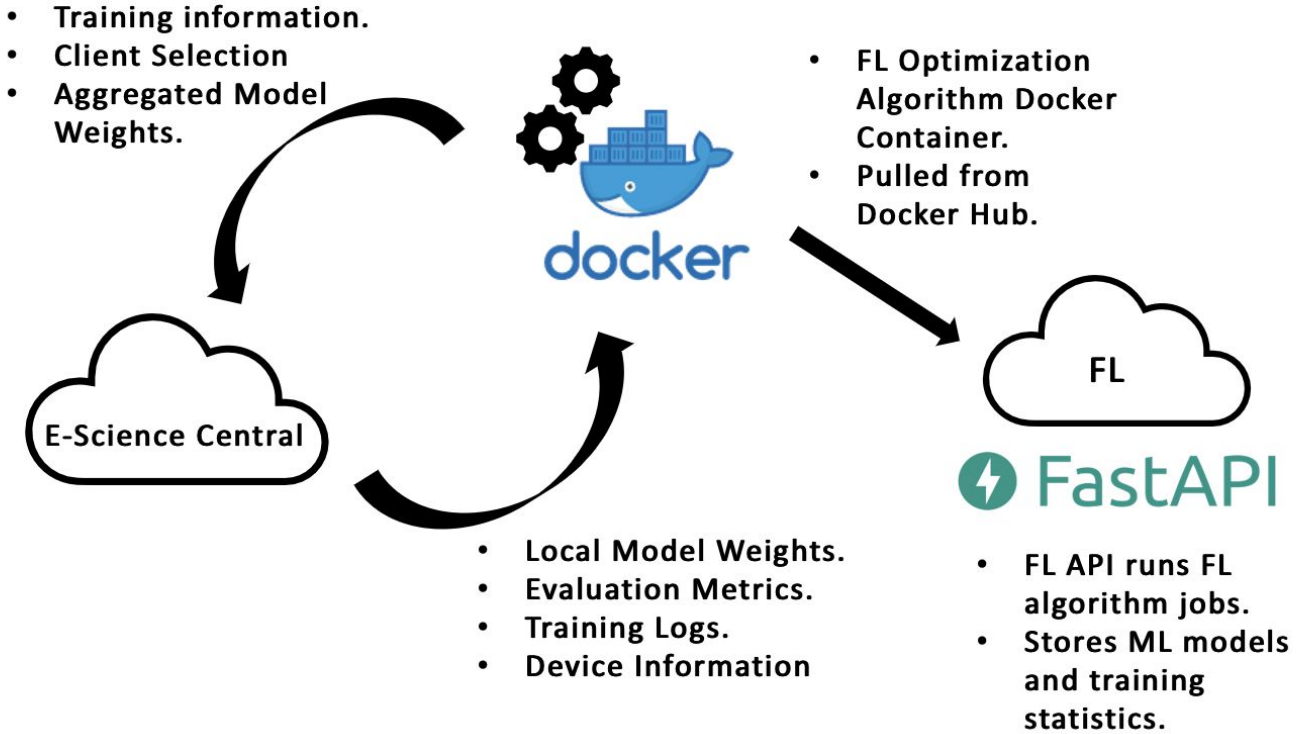
- Handles data processing and storage from wearables.
- Provides dashboard visualizations

- ML docker containers.
- Runs local training
- Pulled from cloud host.

e-Science central

- The newly re-written e-SC is used to manage device registrations and co-ordinate exchange of data files
  - Kubernetes based
  - Focussed on device management, data streaming and edge computing
  - Still in early stages of development
- Edge computing nodes perform local processing
  - Docker containers build local models
- Centralised processing is managed using e-SC Jobs
  - e-SC jobs map directly to Kubernetes jobs
  - The same Docker container structure is used for both centralised and edge processing
- Using e-SC for the low level data management simplifies the model management task

Federated Learning



- Training information.
- Client Selection
- Aggregated Model Weights.

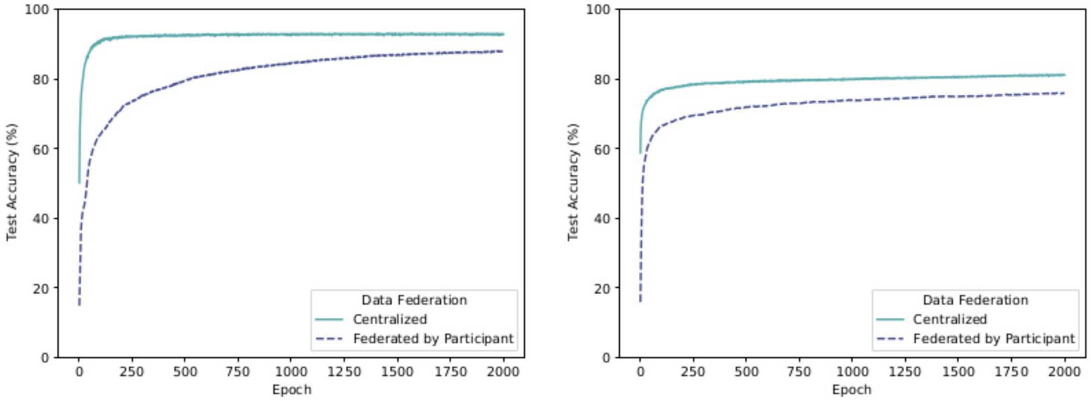
- FL Optimization Algorithm Docker Container.
- Pulled from Docker Hub.

- Local Model Weights.
- Evaluation Metrics.
- Training Logs.
- Device Information

- FL API runs FL algorithm jobs.
- Stores ML models and training statistics.

Simulated Results

- Evaluated on a database of human gait performance on irregular and uneven surfaces collected by wearable sensors.
- Simulated using PySyft.
- Will assess performance using the proposed system.



(a) Deep Neural Network

(b) Logistic Regression (SGD)

Next Stages

- Deployment of the system with Raspberry Pis.
- Assessment of the performance of FedAvg using the gait analytics dataset.
- Improvements on communication and hardware efficiency.
- Deployment of FL algorithms to handle statistical challenges.