Federated Learning of Explainable Artificial Intelligence Models for Predicting Parkinson’s Disease Progression

José Luis Corcuera Bárcena, Pietro Ducange, Francesco Marcelloni, Alessandro Renda, Fabrizio Ruffini

Department of Information Engineering, University of Pisa

IN SUPREMÆ DIGNITATIS · 1343

UNIVERSITÀ DI PISA

Outline

- Towards ***trustworthy*** AI systems
- **Fed-XAI**: ***Fed****erated Learning of* ***XAI*** models
- Case study: progress prediction of Parkinson's Disease in the federated setting
  - Experimental ***setup***: dataset and data distribution scenarios
  - Experimental ***results***: accuracy and interpretability of the Fed-XAI approach

The pursuit of *trustworthiness*

Seven requirements toward Trustworthy AI

- Human agency and oversight
- Accountability
- Technical robustness and safety
- Societal and environmental wellbeing
- Diversity non-discrimination and fairness
- Privacy and data governance
- Transparency

Need to collect (large) data to train accurate ML models clashes with need to preserve privacy of data owners.

"AI systems and their decisions should be explained in a manner adapted to the stakeholder concerned."

Fed-XAI
Federated Learning of eXplainable AI models

DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

IN SUPREMÆ DIGNITATIS 1343

Federated Learning

- A *novel* learning paradigm
  - Training a *centralized model* on *decentralized data*
  - Participants share model updates, not private raw data

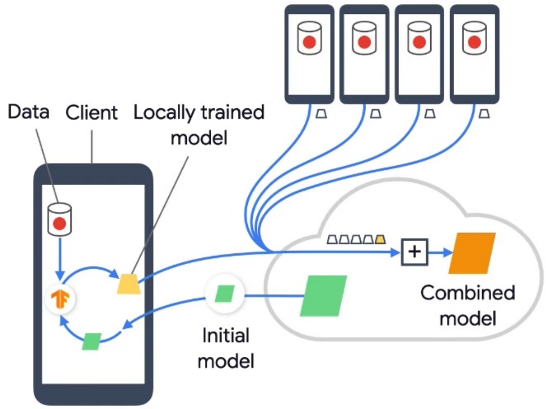
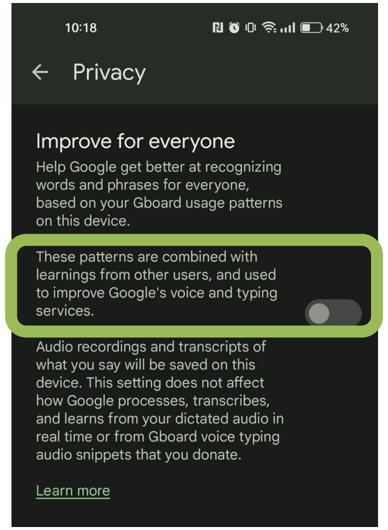


Figure from https://ml.berkeley.edu

- **FedAvg** (iterates over following steps):
  - *server* sends global model to clients
  - *each client* updates the model using local data and sends the model back to the server;
  - *server* takes the average of the locally computed updates, weighted according to the number of samples

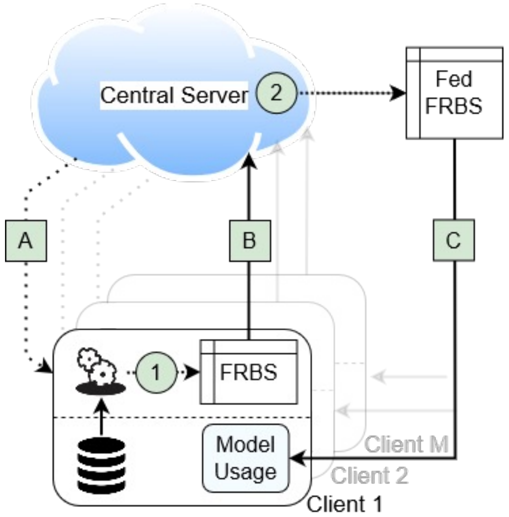


Fed-XAI: Federated Learning of XAI models

- FL is immediately suitable for models in which the learning stage is based on optimization of differentiable global objective function (e.g., DNNs)
- Ad-hoc strategies are needed for **inherently interpretable models,** e.g. **Takagi-Sugeno-Kang (TSK) Fuzzy Rule-Based Systems (FRBS):** collection of rules in the form <*if «antecedent» then «consequent»*>

$$\textbf{IF } X_1 \textbf{ IS } A_{1,j_{k,1}} \ \dots \ \textbf{AND } X_F \textbf{ IS } A_{F,j_{k,F}}$$

$$\textbf{THEN } y_k(\mathbf{x}) = \gamma_{k,0} + \sum_{i=1}^{F} \gamma_{k,i} \cdot x_i$$



A — Configuration: central server sends configuration

1 — Local learning of TSK-FRBSs

B — Transmission of local models to the central server

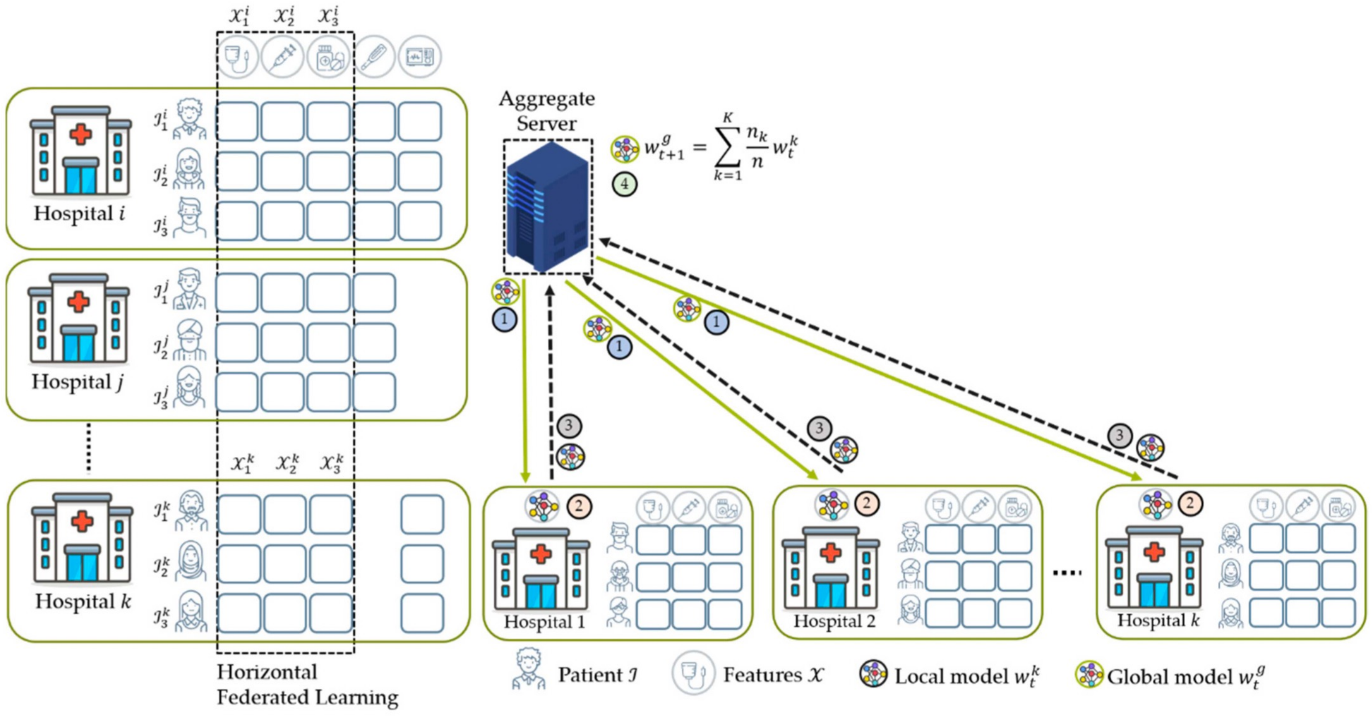
2 — Federated learning of the global TSK-FRBS: aggregation

C — Transmission of the aggregated model to the clients

J. L. Corcuera Bárcena et al. , "An Approach to Federated Learning of Explainable Fuzzy Regression Models," *IEEE Int'l Conf. on Fuzzy Systems (FUZZ-IEEE)*, 2022

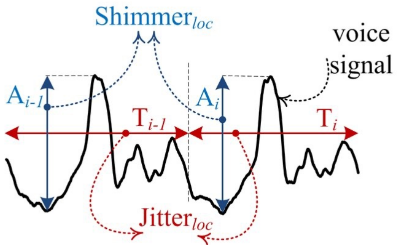
Fed-XAI in healthcare

- Healthcare domain
  - **Privacy preservation** and **explainability** are imperative needs
- Focus on **Horizontal FL**
  - training instances from different hospitals are described by the **same set of features**



- **Objective**
  - Case study: **prediction of Parkinson's Disease progression**
  - Assessing the suitability of the Fed-XAI approach adopting TSK-FRBS as inherently interpretable model

Parkinson’s Disease progression dataset



Input attributes
clinical information and characteristics of voice signal

Target attribute
Unified PD Rating Scale (UPDRS)

PD Telemonitoring Dataset: regression task

- Dataset details
  - 5875 records
  - from 42 subjects (28M,14F)
  - 22 attributes, reduced to 4 through feature selection (*age, test time, Jitter(Abs), DFA*)
  - For TSK-FRBS, each attribute is partitioned with five fuzzy sets (*VeryLow* - *Low* - *Medium* - *High* - *VeryHigh*)

- Federated setting: patients *artificially distributed* into 10 hospitals

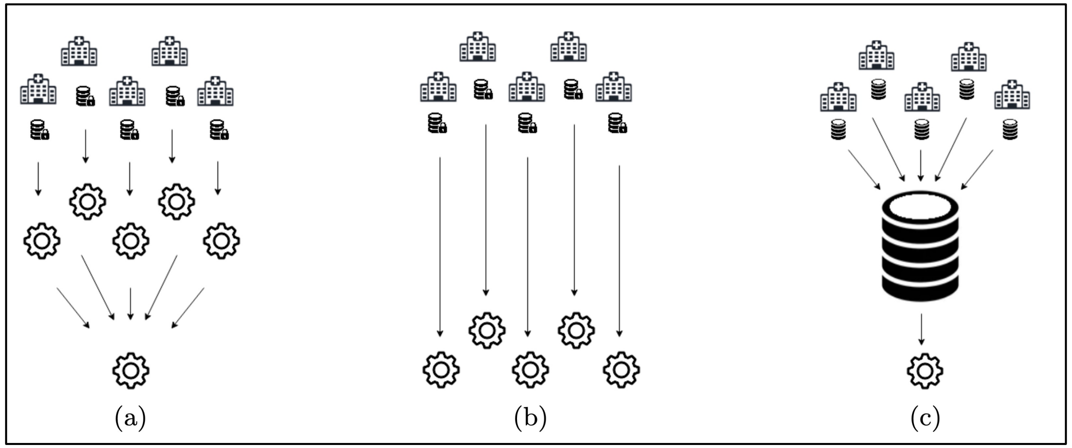
Data distribution scenarios and evaluation strategies

- Scenario 1 (S1): i.i.d. setting: $P_h(\boldsymbol{x}, y) \sim P(\boldsymbol{x}, y) \quad \forall h$, where
  - $P_h(\boldsymbol{x}, y)$ is the local distribution of input $\boldsymbol{x}$ and target y for hospital h
  - $P(\boldsymbol{x}, y)$ is the overall data distribution
- Scenario 2 (S2): non-i.i.d. setting: $P_i(\boldsymbol{x}, y) \neq P_j(\boldsymbol{x}, y)$, for any pair of hospitals (i, j)
  - *Feature distribution skew* for the *age* attribute

	S1: training set		S2: training set		test set	
client	age range	samples	age range	samples	age range	samples
0	[36,85]	529	[36,55]	561	[36,85]	59
1	[36,85]	529	[56,57]	473	[36,85]	59
2	[36,85]	529	[58,59]	655	[36,85]	59
3	[36,85]	529	[60,62]	487	[36,85]	59
4	[36,85]	529	[63,65]	506	[36,85]	59
5	[36,85]	529	[66,66]	380	[36,85]	59
6	[36,85]	529	[67,71]	689	[36,85]	59
7	[36,85]	528	[72,72]	279	[36,85]	59
8	[36,85]	528	[73,74]	591	[36,85]	58
9	[36,85]	528	[75,85]	666	[36,85]	58
tot		5287		5287		588

Experimental evaluation

- (a) Federated Learning (FL)
- (b) Local Learning (LL)
- (c) Centralized Learning (CL)



Experimental results

TSK-FRBS vs **MLP-NN** (*opaque* baseline). **Average values** of the metrics:

- RMSE: *how much predictions deviate from the true values*
- *r* (pearson corr coefficient): *how much predictions and true values are correlated*

General observations

- TSK comparable to MLP, especially in CL
- FL generally outperforms LL both in S1 and S2
  - Noticeable in Scenario 2 (non-i.i.d.), where LL is particularly prone to overfitting
  - Can be appreaciated also in Scenario 1 (i.i.d.), especially for TSK
- FL is generally outperformed by CL
  - Yet unfeasible when privacy preservation is a requirement

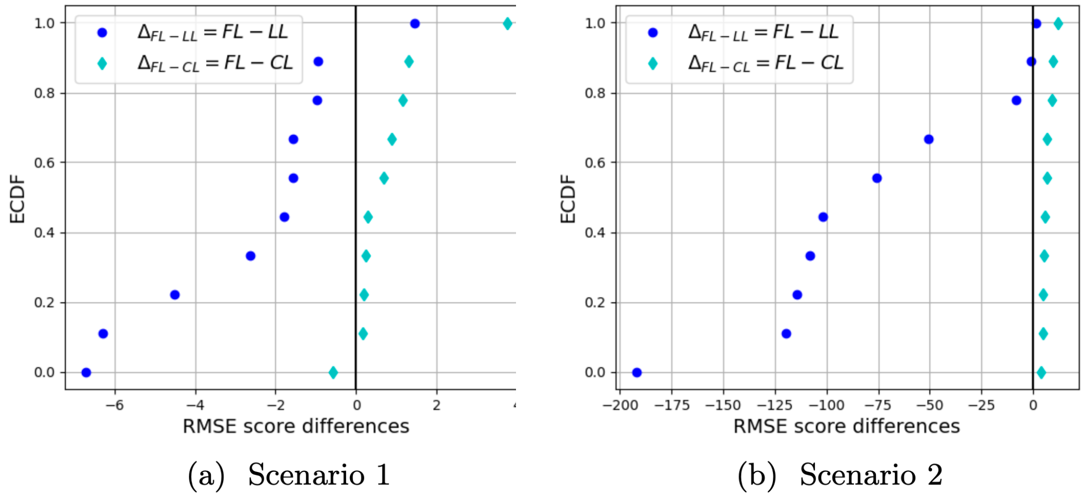
TSK	RMSE		r	
	train	*test*	train	*test*
S1 - LL	6.165	*11.214*	0.820	*0.448*
S1 - FL	7.907	*8.657*	0.677	*0.622*
S1 - CL	7.790	*7.850*	0.688	*0.660*
S2 - LL	3.221	*91.832*	0.919	*-0.064*
S2 - FL	13.166	*14.807*	0.509	*0.470*
S2 - CL	7.477	*7.850*	0.641	*0.660*

MLP	RMSE		r	
	train	*test*	train	*test*
S1 - LL	8.981	*9.122*	0.553	*0.490*
S1 - FL	9.492	*9.192*	0.472	*0.476*
S1 - CL	7.651	*7.722*	0.704	*0.675*
S2 - LL	5.243	*18.108*	0.799	*0.180*
S2 - FL	10.047	*10.150*	0.203	*0.353*
S2 - CL	7.477	*7.657*	0.599	*0.682*

Experimental results: TSK-FRBS

Fine grained analysis (individual hospitals)

- Empirical cumulative distribution function (ECDF) of the differences of RMSE
  - between FL and LL (dark blue)
  - between FL and CL (light blue)
- Pairwise Wilcoxon signed-rank test
  - There is statistical evidence of a difference in performance between FL and LL, and between FL and CL ($\alpha = 0.05$)



Interpretability

- *Global* (structural properties of the model): average number of rules

	LL	FL	CL
S1	217.8	419.0	419.0
S2	71.7	419.0	419.0

- *Local* (inference process)

$$R_k : \textbf{IF} \ age \ is \ VeryHigh \ \textbf{AND} \ test\_time \ is \ Low \ \textbf{AND} \ Jitter(Abs) \ is \ High \ \textbf{AND} \ DFA \ is \ VeryHigh \ \textbf{THEN} : Total\_UPDRS = 0.269 + 0.210 \cdot age + + 0.347 \cdot test\_time + 0.014 \cdot Jitter(Abs) - 0.020 \cdot DFA$$

Conclusions

- **AI in healthcare** poses urgent requirements in terms of **explainability** and **privacy**
- The **Fed-XAI** approach is conceived to simultaneously address the two requirements
- Case study for Parkinson's Disease progression prediction (regression task)
  - Different **data distribution scenarios** for the federated setting
  - Adoption of a **highly interpretable TSK Fuzzy Rule-based System**
  - Comparison with a **MLP-NN as an opaque baseline**

OpenFL-XAI released!

An extension to the open-source OpenFL framework for user-friendly support to Federated Learning of explainable by design models

https://github.com/Unipisa/OpenFL-XAI

Thanks for your attention





alessandro.renda@unipi.it

Artificial Intelligence Group

Dept. of information Engineering
University of Pisa, Italy
Website: **http://ai.dii.unipi.it**

DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

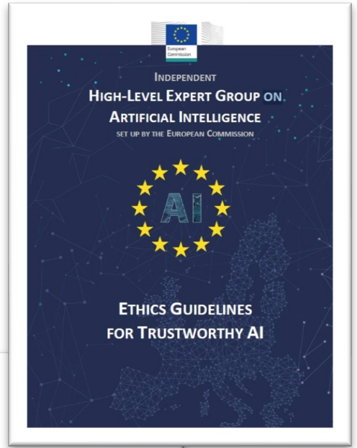
IN SUPREMÆ DIGNITATIS 1343

Backup slides

Trustworthy AI

Citizens and regulators are placing increasing attention on AI trustworthiness

- **AI act** (EU first law on AI):
  - *April 2021* first proposal
  - *May 11, 2023*: press release on draft negotiating mandate
  - *June 15, 2023*: parliament vote for endorsement
- **"Ethic guidelines for trustworthy AI"**, (April 2019)
  - **Lawful:** respecting all applicable laws and regulations
  - **Ethical:** respecting ethical principles and values
  - **Robust:** both from a technical and social perspective

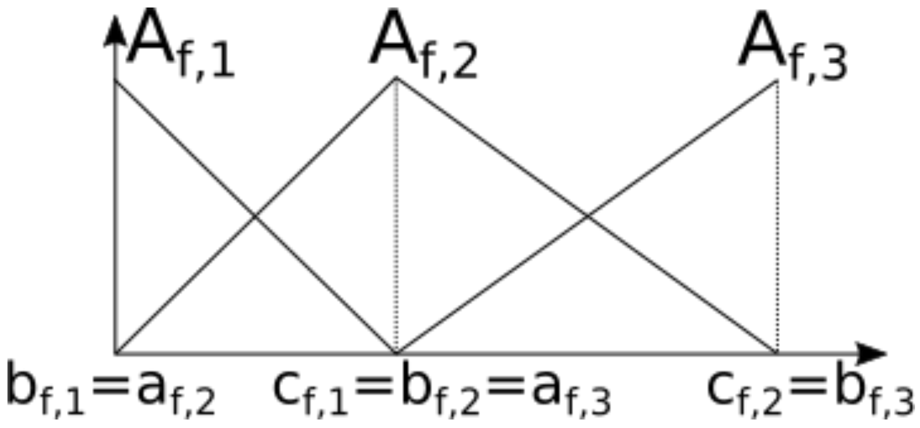


Interpretable Models: Fuzzy Rule-Based Systems

- The model consists of a rule-base, i.e., a collection of rules in the form

  if «antecedent» then «consequent»

- Example of rules in the form *first-order Takagi-Sugeno-Kang Fuzzy Rule-Based Systems*



Strong Fuzzy Partition

$$\textbf{IF } X_1 \textbf{ IS } A_{1,j_{k,1}} \textbf{ ... AND } X_F \textbf{ IS } A_{F,j_{k,F}}$$

$$\textbf{THEN } y_k(\mathbf{x}) = \gamma_{k,0} + \sum_{i=1}^{F} \gamma_{k,i} \cdot x_i$$

Experimental Setup

- **Federated Feature selection**
  - Let $\hat{F}$ be the desired number of features to be selected (we set $\hat{F} = 4$)
  - Preliminary communication round
    - Each client determines the $\hat{F}$ features to be selected based on some importance criterion (RT feature importance - Gini impurity)
    - Each client transmits the candidate list with the server.
    - The server selects the most popular $\hat{F}$ features based on the votes of the clients
    - The server transmits the selected features to the clients

- **MLP-NN hyperparam configuration:**
  - **Architecture**: two hidden layers with 64 neurons each and ReLu as activation function
  - **Loss function**: Mean Squared Error (MSE)
  - **Optimizer**: Adam
  - **FL process parameters**: E=5 (local epochs per round),B=64 (mini-batch size), R=5 (federated rounds)

Case study: Parkinson's disease

Feature name	**Brief description**
subject#	patient identifier
age	Subject age
sex	Subject gender '0' - male, '1' - female
test_time	Time since recruitment into the trial.
motor_UPDRS	Clinician's score, linearly interpolated
total_UPDRS	Clinician's score, linearly interpolated
Jitter[%, Abs, RAP, PPQ5, DDP]	Measures of variation in fundamental frequency
Shimmer, Shimmer[dB, APQ3, APQ5, APQ11, DDA]	Measures of variation in amplitude
NHR, HNR	Two measures of ratio of noise to tonal components in the voice
RPDE	A nonlinear dynamical complexity measure
DFA	Signal fractal scaling exponent
PPE	A nonlinear measure of fundamental frequency variation