Understanding Clinical Collaborations Through Federated Classifier Selection

Sebastian Caldas[1], Joo H. Yoon[2], Michael R. Pinsky[2], Gilles Clermont[2], Artur Dubrawski[1]

[1]Auton Lab, School of Computer Science, Carnegie Mellon University
[2]School of Medicine, University of Pittsburgh

Contributions

- We argue for the importance of of understanding how a collaboration may be affecting the quality of a clinical center's predictions.
- We propose FRCLS, an algorithm that finds regions of the feature space where external models outperform the local model, and describes these regions of expertise through simple rules.
- We demonstrate the effectiveness of FRCLS on two different hospital systems in the context of an early sepsis prediction task.

Motivation

- Previous work in federated learning for healthcare has equated utility with predictive power, neglecting other aspects of clinical utility.
- We are interested in explaining how a clinical collaboration itself is affecting a center's predictions, e.g., whether a decision is being made based on knowledge from an external center.
- Rationale of this type can incentivize further cooperation, inform local resource allocation, or even help identify external best practices.

FedeRated CLassifier Selection (FRCLS)

FRCLS proceeds in three stages:

1. Training of local classifiers.
2. Exchange of classifiers.
   - Each hospital is left with a local classifier c_L and a pool of external classifiers $\{c_m\}_{m=1}^{M}$.
3. Dynamic selection of candidate classifiers.
   - Happens independently at each center.

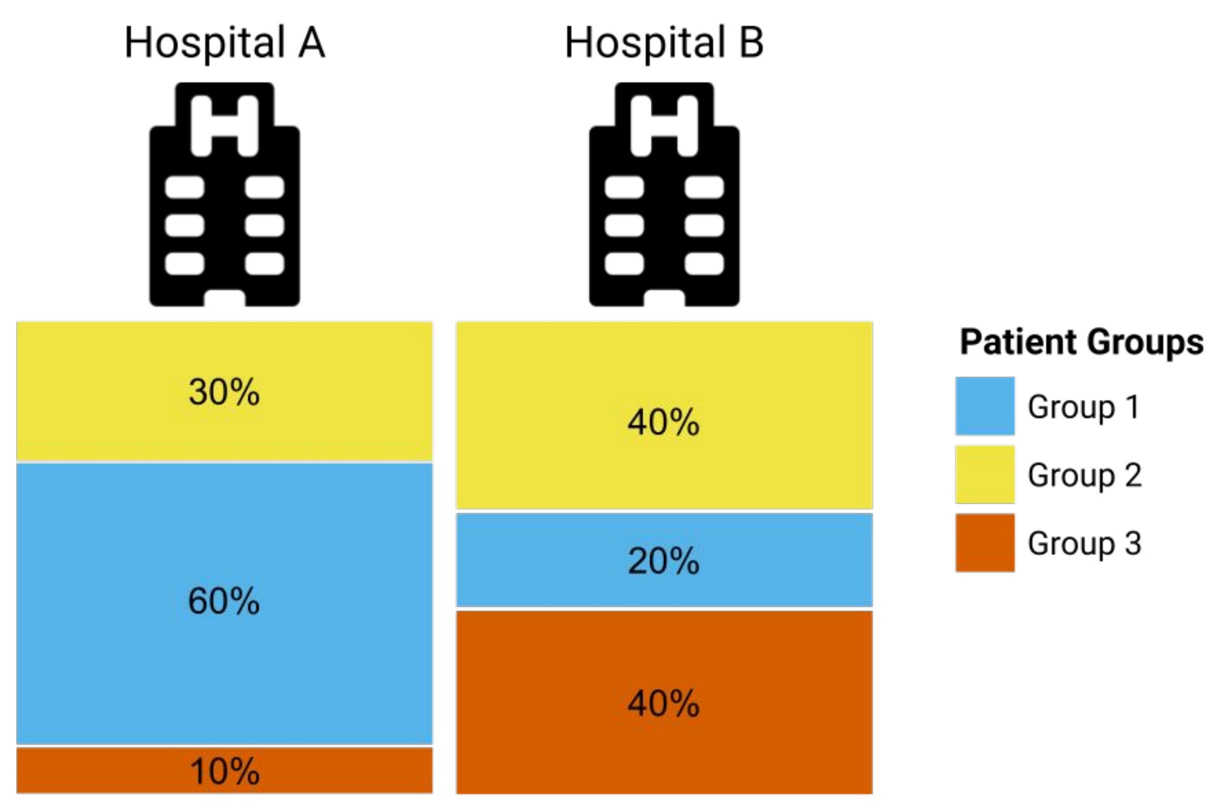

Figure 1: Intuition behind FRCLS. Inter-center population heterogeneity makes each hospital an expert on different patient subpopulations. FRCLS leverages this diversity among classifiers and dynamically picks the model that is best for each incoming instance.

References:
[1] M. Reyna et al. Early prediction of sepsis from clinical data: the physionet/computing in cardiology challenge, *CinC*, 2019.

Dynamic Selection of Candidate Classifiers

- For each new instance x, we wish to determine whether to use:
  - The local classifier c_L.
  - A greedy external classifier $c_E(x) = \arg\max_{c_m \in C} c_m(x)$
- Define

$$L_c(x,k) = \frac{1}{k} \sum_{j \in nn(x,k)} \ell(c(x_j), y_j)$$

$$\rho_E(x,k) = \frac{L_{c_L}(x,k)}{L_{c_E}(x,k)}$$

- We use c_E if $\rho_E(x) > \rho_0$ where ρ_0 minimizes the p-value of the test:

$$F(\rho_0) = |\{x_i : \rho_E(x_i,k) > \rho_0, c_L^*(x_i) \neq c_E^*(x_i), c_E^*(x_i) \neq y_i\}|$$
$$S(\rho_0) = |\{x_i : \rho_E(x_i,k) > \rho_0, c_L^*(x_i) \neq c_E^*(x_i).c_E^*(x_i) = y_i\}|$$
$$H_0 : \frac{S(\rho_0)}{S(\rho_0) + F(\rho_0)} < 0.5$$

- A second strategy uses a rule learning algorithm to create a decision list that maximizes a lower bound on the mean of ρ_E. We use c_E if x satisfies the rules.

Results and Discussion

- We demonstrate our method on the early sepsis prediction task proposed by [1].
- The data corresponds to ICUs in two hospital systems. We call them A and B.
- Our local classifiers are logistic regression models with ridge penalty.

Hospital System	Val p-value	p-value	Instances handled by c_E	Successful Flips (% of flips)	Local Accuracy	External Accuracy
A (@90% TPR)	**2.52e−1**	-	0	-	-	-
A (@10% FPR)	1.12e−6	3.07e−5	1925	299 (58.97%)	57.92%	62.65%
B (@90% TPR)	1.35e−3	1.25e−8	700	134 (70.16%)	51.43%	62.43%
B (@10% FPR)	**8.04e−2**	-	0	-	-	-

Figure 2: Results for our decision list strategy. When the p-value on the validation set is greater than 0.05 (bolded), no instances are handled by c_E. Accuracies are given for those instances where FRCLS uses c_E over c_L.

Hospital System A

```
IF PTT > 84.55 AND Phosphate <= 8.42:
    Use c_E
ELIF BUN > 83.56 AND Calcium <= 9.66:
    Use c_E
ELIF Hct > 40.31:
    Use c_E
ELIF Calcium > 9.66:
    Use c_E
ELIF Hgb > 12.14:
    Use c_E
ELSE:
    Use c_L
```

Hospital System B

```
IF BaseExcess > -0.92:
    Use c_E
ELIF FiO2 > 0.65:
    Use c_E
ELSE:
    Use c_L
```

Figure 3: Rules learned by FRCLS's decision list strategy.