Preserving Patient Privacy while Training a Predictive Model of In-Hospital Mortality

Pulkit Sharma, Farah E. Shamout, David A. Clifton
Department of Engineering Science, University of Oxford, UK
pulkit.sharma@eng.ox.ac.uk, farah.shamout@eng.ox.ac.uk, davidc@robots.ox.ac.uk

Introduction 1

- Patient data is often collected at different hospitals and sharing is restricted due to patient privacy concerns.
- Deep learning typically requires large quantities of training data to learn complex models.
- We discuss the potential of distributed training in achieving state-of-the-art performance while maintaining data privacy.
- The model is trained in a federated learning framework which leads to comparable performance to the traditional centralised setting.

Framework 2

- Assume a set of hospitals $\mathcal{H} = \{\mathcal{H}_1, \ldots, \mathcal{H}_K\}$, with a common server S coordinating between them.
- Each hospital $\mathcal{H}_k$ stores its data $D_k = \{(x_1^k, y_1^k), (x_2^k, y_2^k), \ldots (x_{|D_k|}^k, y_{|D_k|}^k)\}$ locally.
- x_i^k and y_i^k represent the data sample i and its corresponding label, respectively, at hospital k.
- $|D_k|$ represents the total number of data samples stored at hospital k.

Federated Learning (FL) 3

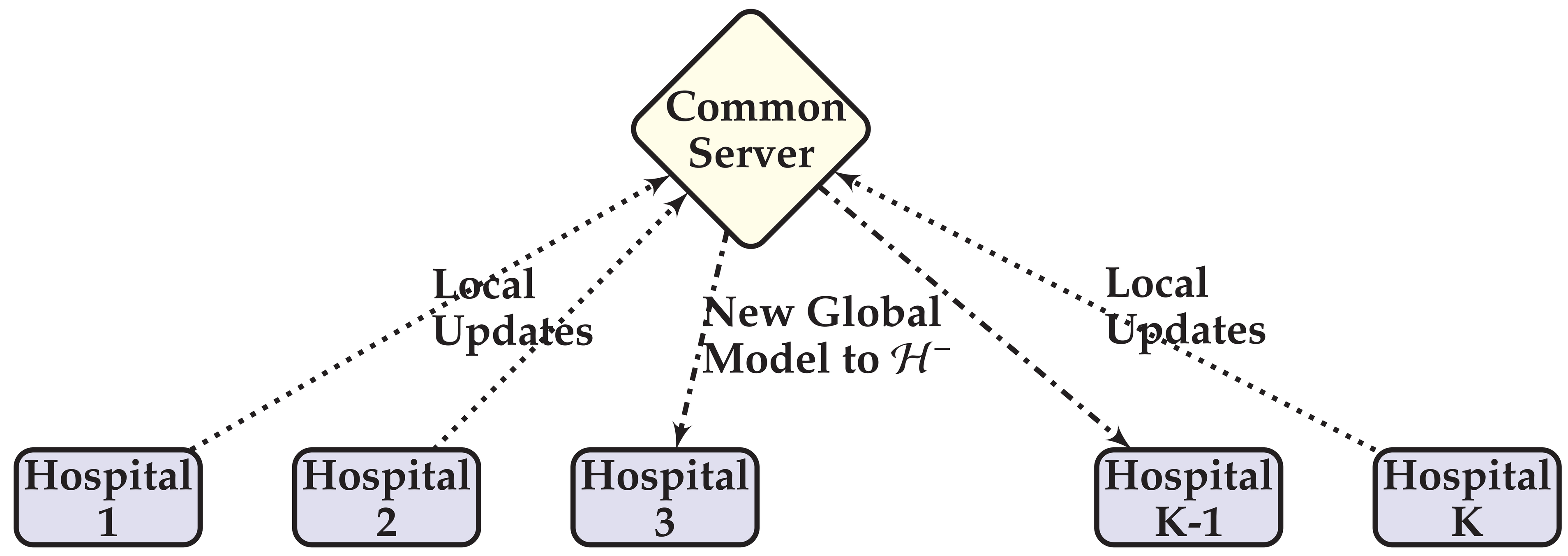
- The goal is to estimate the global (G) parameters $\mathbf{w}^G \in \mathbb{R}^d$ of the global model without directly accessing the data stored at the hospitals, where d represents the number of parameters of the model.
- S broadcasts the global model $\mathbf{w}_t^G$ to a subset of non-identically-distributed hospitals $\mathcal{H}^- \subset \mathcal{H}$ at time t.
- A local loss is optimised over the local data D_k at every node k in $\mathcal{H}^-$ to estimate the local parameter vector $\mathbf{w}_{t+1}^k$.
- Hospitals $\mathcal{H}^-$ send their computed model parameters S, which aggregates the findings to estimate an updated global model $\mathbf{w}_{t+1}^G$ as:

$$\mathbf{w}_{t+1}^G = \sum_{k=1}^{|\mathcal{H}^-|} p_{t+1}^k \mathbf{w}_{t+1}^k. \quad (1)$$

FL: Aggregation 4

- Dropping the time dimension (for simplicity) we consider one time instance $\mathbf{w}^G = \sum_{k=1}^{|\mathcal{H}^-|} p^k \mathbf{w}^k$.
- $p^k \in [0, 1]$ represents the weights associated with each hospital k such that $\sum_{k=1}^{|\mathcal{H}^-|} p^k = 1$.
- Initialise $p^k = \frac{|D_k|}{D}$, where $D = \sum_k |D_k|$ is the total number of samples across the k hospitals.
- Iterate through the described training procedure across different hospitals until convergence or some stopping criterion.
- At each step, the model can be updated locally at each hospital in $k \in \mathcal{H}^-$.
- The model is evaluated using the test data at all the hospitals; i.e. $k \in \mathcal{H}$.
- Accuracy a_t (on held-out test data $k \in \mathcal{H}$) is used as metric of evaluation to update the global model, where a model is updated only if $a_{t+1} \geq a_t$.

Problem Formulation and Proposed Algorithm 5



Schematic of the federated learning (FL) framework adopted for the in-hospital mortality prediction task. In order to preserve the privacy of clinical data, the model is trained in a distributed fashion: The hospitals periodically communicate the local updates with a common server to learn a global model. The common server incorporates the updates and sends back the parameters of the updated global model.

Algorithm 1 A summary of the FL framework to compute the global model at common server using data stored locally at different hospitals. Functions ModelUpdate and LocalTestAccuracy are executed locally on the k^{th} hospital. Variable a_t is an estimation of the global accuracy at time t.

Input: $\mathbf{w}_t^G$ a_t
Output: $\mathbf{w}_{t+1}^G$ a_{t+1}

```
broadcast w_t^G to hospitals in H^-
for each hospital k ∈ H^- do
    w_{t+1}^k ← ModelUpdate(k, w_t^G)
    p_{t+1}^k ← |D_k| / Σ_k |D_k|
end for
w̃_{t+1}^G ← Σ_{k=1}^{|H^-|} p_{t+1}^k w_{t+1}^k
for each hospital k ∈ H do
    a_{t+1}^k ← LocalTestAccuracy(k, w̃_{t+1}^G)
end for
a_{t+1} ← weighted average of a_{t+1}^k ∀ k ∈ H
while a_{t+1} < a_t
    w̃_{t+1}^G ← w_t^G
    a_{t+1} ← a_t
end while
w_{t+1}^G ← w̃_{t+1}^G
```

Experimental Setup 6

- The proposed FL framework is evaluated for the task of predicting in-hospital mortality using the MIMIC-III database [1].
- The total number of patient admissions for the mortality prediction task is 21,138, where the variables collected in the first 48-hour window are used as input features [2].
- The number of time-stamped observations seen in the first 48 hours varied per patient episode. Hence, we used hand-engineered features as described in [3].
- To mimic the FL framework described, we distributed the training and testing data amongst virtual workers [4].

Results 7

Comparison of the proposed FL methods with the standard setup. LR-ORG/MLP-ORG and LR-FL/MLP-FL represents logistic regression/multi-layer perceptron classifier trained in normal and FL configuration.

	LR-ORG	LR-FL
AUROC	0.8152	0.7890
AUPRC	0.4030	0.3659

	MLP-ORG	MLP-FL
AUROC	0.7925	0.7769
AUPRC		0.3900
0.3504		

Conclusions 8

- FL-based models perform well for the in-hospital mortality prediction task, while preserving patient privacy.
- With improved data privacy, data owners would be more comfortable in utilising their data for machine learning research by not sharing the data directly.
- FL may allow us to train machine learning models on larger and potentially more diverse datasets, which would also improve the performance.

References 9

[1] A.E.W. Johnson, T.J. Pollard, L. Shen, L.H. Lehman, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L.A. Celi, and R.G. Mark. MIMIC-III, a freely accessible critical care database. Scientific Data, 3, 2016.
[2] H. Harutyunyan, H. Khachatrian, D.C. Kale, G.V. Steeg, and A. Galstyan. Multitask learning and benchmarking with clinical time series data. Scientific Data, 6, 2019.
[3] Z.C. Lipton, D.C. Kale, C. Elkan, and R.C. Wetzel. Learning to diagnose with LSTM recurrent neural networks. In International Conference on Learning Representations (ICLR), 2016.
[4] coMindOrg. comindorg/federated-averaging-tutorials, Mar 2019.