

Title. Development of a federated learning approach to predict acute kidney injury in adult hospitalized patients with COVID-19 in New York City

Author Block. Suraj K Jaladanki^{1,2+}, BS; Akhil Vaid^{1,2+}, MD; Ashwin S Sawant^{2,3}, MD; Jie Xu⁴, PhD; Kush Shah², BA; Sergio Dellepiane^{2,3}, MD, PhD; Ishan Paranjpe^{1,2}, BS; Lili Chan^{3,5,6,7}, MD, MS; Patricia Kovatch⁸, BS; Alexander W Charney^{2,8,9,10}, MD, PhD; Fei Wang⁴, PhD; Benjamin S Glicksberg^{1,2,8}, PhD; Karandeep Singh¹¹, MD, MS; Girish N Nadkarni^{1,2,3,5,7*}, MD, MPH

+ Equal contribution

* Corresponding author

Running Title: Federated Learning for AKI Prediction in COVID-19

Affiliations.

1. The Hasso Plattner Institute for Digital Health at Mount Sinai, Icahn School of Medicine at Mount Sinai, New York, New York, USA
2. The Mount Sinai Clinical Intelligence Center (MSCIC), New York, New York, USA
3. The Division of Data-Driven and Digital Medicine (D3M), Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, New York, USA
4. Department of Population Health Sciences, Weill Cornell Medicine, New York, New York, USA

5. The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, New York, USA
6. BioMe Phenomics Center, Icahn School of Medicine at Mount Sinai, New York, New York, USA
7. Division of Nephrology, Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, New York, USA
8. Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, New York, USA
9. The Pamela Sklar Division of Psychiatric Genomics, Icahn School of Medicine at Mount Sinai, New York, New York, USA
10. Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, New York, USA
11. Department of Learning Health Sciences, University of Michigan Medical School, Ann Arbor, Michigan, USA

Correspondence:

Dr. Girish N. Nadkarni
Icahn School of Medicine at Mount Sinai
One Gustave L. Levy Place, Box 1003
New York, NY 10029
girish.nadkarni@mountsinai.org
T: 212-241-1385
F: 212-241-1385

Sources of support. This study was supported by the National Center for Advancing Translational Sciences, National Institutes of Health (U54 TR001433-05). GNN is also supported by R01DK127139, R01HL155915, and R56DK126930. FW would like to acknowledge the support from NSF 1750326 and NIH RF1AG072449.

ABSTRACT

Federated learning is a technique for training predictive models without sharing patient-level data, thus maintaining data security while allowing inter-institutional collaboration. We used federated learning to predict acute kidney injury within three and seven days of admission, using demographics, comorbidities, vital signs, and laboratory values, in 4029 adults hospitalized with COVID-19 at five sociodemographically diverse New York City hospitals, between March-October 2020. Prediction performance of federated models was generally higher than single-hospital models and was comparable to pooled-data models. In the first use-case in kidney disease, federated learning improved prediction of a common complication of COVID-19, while preserving data privacy.

Keywords: Federated learning; COVID-19; Acute Kidney Injury; machine learning; privacy protection; electronic health records

Introduction

Machine Learning (ML) models have been developed to aid in risk-stratification in patients with Coronavirus Disease-19 (COVID-19).¹⁻³ Due to concerns about data privacy and ownership, most of these models were trained on data from single institutions, which limits their generalizability.^{4, 5} However, it is critical that these models be trained on large, diverse datasets, reflecting the variability seen in clinical practice.

Federated learning (FL), a technique that allows model training to occur without sharing confidential patient data, has gained significant interest during the pandemic. FL avoids the need to collect and store private data at a centralized location by allowing institutions to download a preliminary ML model, refine it locally, and then upload updated model parameters to an aggregator.⁵⁻⁷ FL models have been utilized to identify COVID-19 through computed tomography scans and to predict clinical outcomes in COVID-19 patients — published reports indicate that FL models not only outperform those trained on single-institution data but also approach the performance of traditional ML models trained on data pooled from multiple institutions.^{6, 8-10}

COVID-19 has diverse clinical manifestations, and a common complication in hospitalized COVID-19 patients is acute kidney injury (AKI).¹¹⁻¹⁴ Studies have reported AKI prevalence up to 46%, and mortality rates in AKI cohorts vary from 35%–71%.^{15, 16} Clinical models to pre-emptively identify patients with COVID-19 at increased risk for developing AKI can be valuable during a pandemic when resources may be constrained.¹⁷

Using COVID-19 associated AKI as a use case, we aimed to evaluate how FL trained models compare to models trained at individual institutions and pooled multi-institutional data. We developed federated models to predict AKI within three and seven days of admission using electronic health records (EHR) in patients hospitalized with COVID-19 and compared their performance to local and pooled models.

Short Methods

The study population consisted of adult patients admitted to one of five Mount Sinai Health System hospitals in New York City with laboratory confirmed SARS-CoV-2 infection within 48 hours of admission between March 1st and October 18th, 2020. We excluded patients with a history of transplants, a diagnosis of kidney failure or admissions less than 48 hours. We defined AKI according to 2012 Kidney Disease Improving Global Outcomes (KDIGO) guidelines.¹⁸ The primary outcomes of interest were incident AKI within three (AKI₃) and seven (AKI₇) days of admission. We included demographics, comorbidities, vital signs, laboratory values, and clinical outcomes extracted from an EHR database. For each parameter, we used the first available values obtained within 48 hours after admission. We used two classifiers: multilayer perceptron (MLP) to represent deep learning and Least Absolute Shrinkage and Selection Operator (LASSO) as a regression method. We used three training strategies for each classifier: local, federated, and pooled (Figure 1A, 1B). Local models for each hospital were trained and tested using only data from that hospital. Federated models used a FL framework to share model parameters with a central aggregator. Finally, pooled models combined data from all hospitals for model development and

represented an optimal scenario. Model performance was assessed by average areas under the receiver operating characteristic curve (AUROCs) across 100 independent experiments with 70%/30% training/testing splits. SHapley Additive exPlanations (SHAP) scores were used to evaluate the importance of features contributing to model predictions.¹⁹ Detailed inclusion and exclusion criteria, data processing, classifier parameters, and software are described in Detailed Methods (Supplementary Material). We used the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) guidelines and shared our code under the GNU General Public License v3 to enhance replicability (Supplementary Material).

Results

Cohort Characteristics

Among 4029 patients, the prevalence of AKI₃ ranged from 23% to 37% across five hospitals, and 27% to 44% for AKI₇. There were significant differences in demographics, comorbidities, laboratory measurements, outcome prevalence, and sample sizes between hospitals (Supplementary Table S3). For example, Mount Sinai West (MSW) had the largest class imbalance and smallest sample size of all hospitals.

Learning Framework Comparisons

We evaluated performance of three model strategies (local, pooled, and federated) at each hospital for predicting AKI₃ and AKI₇. Performance averaged across the five sites is shown in Figure 1C. All models had AUROCs in the range of 0.70–0.90 for predicting AKI₃ and AKI₇. Pooled models consistently outperformed both local and

federated models for AKI₃ and AKI₇ prediction. LASSO_{federated} also outperformed LASSO_{local} at all hospitals except one for both AKI₃ and AKI₇ prediction (Supplementary Table S5). MLP_{federated} outperformed MLP_{local} only at MSW for both AKI₃ and AKI₇ prediction (Supplementary Table S5). Additional performance metrics for all models are available in the supplementary material (Supplementary Table S4).

MSW Performance and Feature Importance

The largest improvements in performance between local and federated models were seen for MSW (Figure 2A), with an increase of at least 3% (0.03) in AUROC for all three classifiers for AKI₃ and AKI₇ prediction. SHAP plots for LASSO_{local} and LASSO_{federated} models for AKI₃ prediction at MSW demonstrate differences in feature importance (Figure 2B). For LASSO_{local}, history of stroke, black race, and Hispanic/Latino ethnicity were the three most important features for predicting AKI₃. However, for the corresponding LASSO_{federated} model, the three features with highest importance were blood urea nitrogen, age, and albumin.

Discussion

We trained models using federated learning to predict AKI at three and seven days in hospitalized patients with COVID-19. To our knowledge, this is the first use case of federated learning in kidney disease and to predict AKI over multiple time horizons. We found that models trained with FL outperformed those trained on locally available data, which was most apparent at the hospital with the smallest dataset.

All models had AUROCs of at least 0.70, suggesting that early vital signs and laboratory measurements can predict kidney complications during the first week in patients hospitalized with COVID-19. A similar approach has been used to predict mortality in hospitalized patients with COVID-19, highlighting the utility of peri-admission measurements in the prediction of clinical outcomes in this cohort.¹⁰ Model performance was consistently lower for AKI₇ than AKI₃, likely due to the longer prediction interval in the former case.

We found that LASSO_{federated} models outperformed their respective LASSO_{local} models at most hospitals. At MSW, the hospital with the smallest sample size and largest class imbalance, SHAP plots indicate that the LASSO_{federated} model attached greater importance to features that have more biologically plausible links to renal outcomes than the corresponding LASSO_{local} model, which was strongly influenced by race and a history of stroke. MSW was also the only site where MLP_{federated} outperformed MLP_{local}, indicating that the greatest benefits from using FL to train deep learning models may accrue to sites with small amounts of local data available for training.

Models developed during this study may be implemented in clinical practice to predict onset of AKI in COVID-19 patients several days before it occurs, allowing clinicians to triage patients for better monitoring and use of intravenous fluids. Early prediction of AKI also allows for proactive allocation of resources, because published data indicate that these patients are hospitalized longer, are more likely to need intensive care, and more likely to require inpatient dialysis.¹⁴

Our study has several limitations. First, our data was limited to a single hospital system within New York City, which may limit generalizability to other geographical regions. Similarly, as the standard of care for hospitalized patients with COVID-19 has evolved after our data was collected, model performance should be validated with newer cohorts. Our models did not use inputs such as imaging and echocardiograms, and their exclusion may have hindered performance. We deliberately limited hyperparameter tuning of models to simulate a scenario where a federated model has to be urgently deployed at resource-constrained hospitals, such as during the COVID-19 pandemic. Allowing further optimization of the MLP structure, and techniques such as transfer learning may improve $MLP_{\text{federated}}$ model performance.²⁰ We examined only two classifiers; other approaches such as random forests or support vector machines may have yielded better results. Lastly, performance with the FL approach was generally statistically inferior to pooled dataset performance, and it remains to be explored if other approaches exist which can help bridge this gap while retaining the privacy advantages of FL.

In summary, we demonstrate the utility of FL to improve prediction of key outcomes while maintaining privacy and confidentiality. We hope this will encourage the development of generalizable clinical models which would otherwise be hindered by inability to share patient-level data across institutional boundaries.

References

1. Mei X, Lee H-C, Diao K-y, *et al.* Artificial intelligence-enabled rapid diagnosis of patients with COVID-19. *Nature Medicine* 2020; **26**: 1224-1228.
2. Vaid A, Somani S, Russak AJ, *et al.* Machine Learning to Predict Mortality and Critical Events in a Cohort of Patients With COVID-19 in New York City: Model Development and Validation. *J Med Internet Res* 2020; **22**: e24018.
3. Wynants L, Van Calster B, Collins GS, *et al.* Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. *BMJ* 2020; **369**: m1328.
4. Cahan EM, Hernandez-Boussard T, Thadaney-Israni S, *et al.* Putting the data before the algorithm in big data addressing personalized healthcare. *npj Digital Medicine* 2019; **2**: 78.
5. Song X, Yu ASL, Kellum JA, *et al.* Cross-site transportability of an explainable artificial intelligence model for acute kidney injury prediction. *Nature Communications* 2020; **11**: 5668.
6. Flores M, Dayan I, Roth H, *et al.* Federated Learning used for predicting outcomes in SARS-COV-2 patients. *Res Sq* 2021.
7. Xu J, Glicksberg BS, Su C, *et al.* Federated learning for healthcare informatics. *arXiv preprint* 2020.
8. Kumar R, Khan AA, Zhang S, *et al.* Blockchain-Federated-Learning and Deep Learning Models for COVID-19 detection using CT Imaging. *arXiv preprint* 2020.
9. Xu Y, Ma L, Yang F, *et al.* A collaborative online AI engine for CT-based COVID-19 diagnosis. *medRxiv* 2020: 2020.2005.2010.20096073.
10. Vaid A, Jaladanki SK, Xu J, *et al.* Federated Learning of Electronic Health Records to Improve Mortality Prediction in Hospitalized Patients With COVID-19: Machine Learning Approach. *JMIR Med Inform* 2021; **9**: e24207.
11. Wu Z, McGoogan JM. Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72 314 Cases From the Chinese Center for Disease Control and Prevention. *JAMA* 2020; **323**: 1239-1242.

12. Chung M, Bernheim A, Mei X, *et al.* CT Imaging Features of 2019 Novel Coronavirus (2019-nCoV). *Radiology* 2020; **295**: 202-207.
13. Lala A, Johnson KW, Januzzi JL, *et al.* Prevalence and Impact of Myocardial Injury in Patients Hospitalized With COVID-19 Infection. *J Am Coll Cardiol* 2020; **76**: 533-546.
14. Chan L, Chaudhary K, Saha A, *et al.* AKI in Hospitalized Patients with COVID-19. *Journal of the American Society of Nephrology* 2021; **32**: 151-160.
15. Zahid U, Ramachandran P, Spitalewitz S, *et al.* Acute Kidney Injury in COVID-19 Patients: An Inner City Hospital Experience and Policy Implications. *American Journal of Nephrology* 2020; **51**: 786-796.
16. Hirsch JS, Ng JH, Ross DW, *et al.* Acute kidney injury in patients hospitalized with COVID-19. *Kidney International* 2020; **98**: 209-218.
17. Goldfarb DS, Benstein JA, Zhdanova O, *et al.* Impending Shortages of Kidney Replacement Therapy for COVID-19 Patients. *Clinical Journal of the American Society of Nephrology* 2020; **15**: 880-882.
18. Kidney Disease: Improving Global Outcomes (KDIGO) Acute Kidney Injury Work Group. KDIGO Clinical Practice Guidelines for Acute Kidney Injury. *Kidney inter., Suppl.* 2012; **2**: 1-138.
19. Lundberg SM, Erion G, Chen H, *et al.* From Local Explanations to Global Understanding with Explainable AI for Trees. *Nature machine intelligence* 2020; **2**: 56-67.
20. Ching T, Himmelstein DS, Beaulieu-Jones BK, *et al.* Opportunities and obstacles for deep learning in biology and medicine. *J R Soc Interface* 2018; **15**: 20170387.

Figure Legends

Figure 1. Study Design and Model Performance.

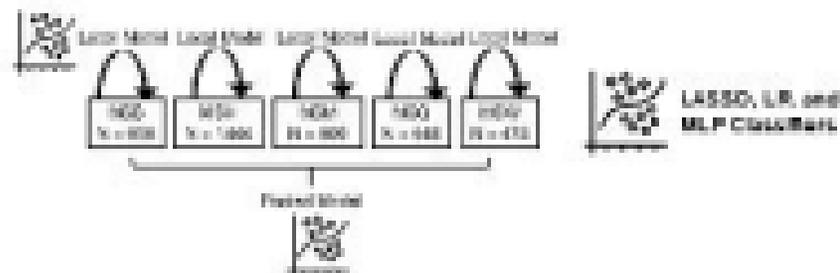
(A) Overview of local and pooled models. Local models only utilize data from the site itself while pooled models incorporate data from all sites. Both local and pooled MLP, LR, and LASSO models were utilized. Based on <https://medinform.jmir.org/2021/1/e24207/> (B) Overview of federated model. Parameters from a central aggregator are shared with each site, and sites do not have direct access to clinical data from others. After models are trained locally at a site, parameters are sent back to the central aggregator to update federated model parameters. Federated LASSO and MLP models were utilized. Based on <https://medinform.jmir.org/2021/1/e24207/> (C) Performance of all models averaged across all sites as measured by area under the receiver-operating characteristic (AUROC) after 70-30 train-test split over 100 experiments with 95% confidence intervals for predicting AKI within three (top) and seven (bottom) days of admission.

Figure 2. Model Performance and Feature Importance at MSW

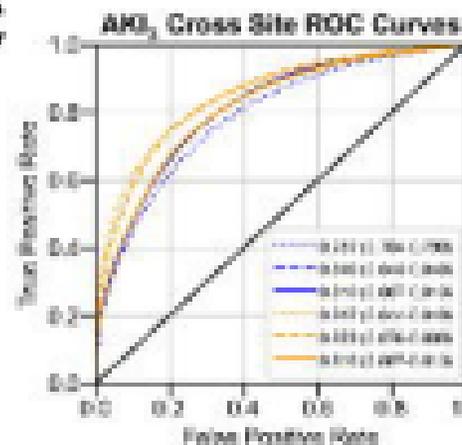
(A) Performance of all models at Mount Sinai West (MSW) (n=474) as measured by area under the receiver-operating characteristic (AUROC) after 70-30 train-test split over 100 experiments with 95% confidence intervals for predicting AKI within three (left) and seven (right) days of admission. (B) SHAP values were calculated for LASSO_{local} (left) and LASSO_{federated} (right) for predicting AKI within three days of admission at MSW and illustrated in summary plots where features are listed in decreasing order of importance.

A

Local and Pooled Models

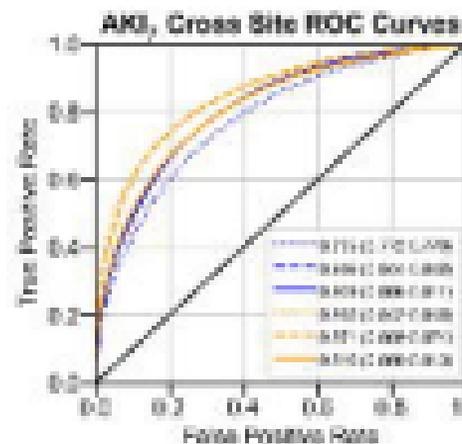
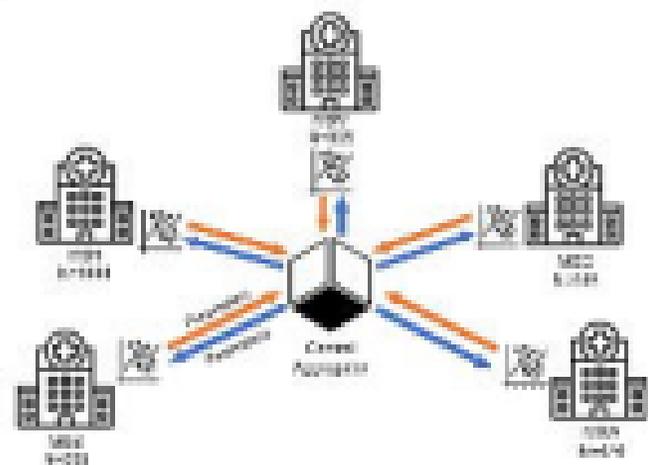


C

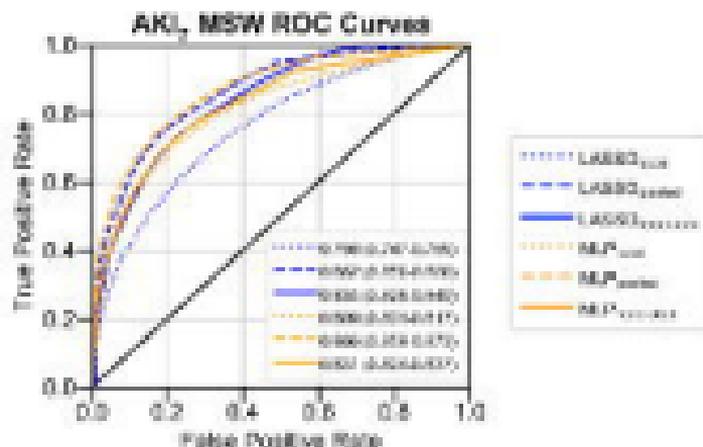
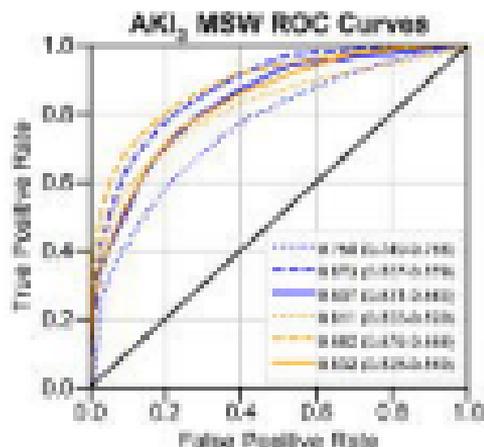


B

Federated Model



A



B

