FEDERATED LEARNING FRAMEWORK FOR NLP IN HEALTHCARE: ASSESSING HOSPITAL READMISSION USING ELECTRONIC HEALTH RECORDS

Bio-IT World, Boston, May 15-18, 2023

Federated Learning Framework for NLP in Healthcare: Assessing Hospital Readmission Using Electronic Health Records

Sahil Nalawade[1], MS, Soujanya Samineni[1], MS, Alex Chowdhury[1], MS, Ling Feng[1], MS, Renato Umeton[1,2,3,4], PhD, Michael Rosenthal[5-7], MD, PhD

1 AI Operations and Data Science Services, Dana-Farber Cancer Institute, Boston, MA, 2 Massachusetts Institute of Technology, Boston, MA, 3 Harvard T.H. Chan School of Public Health, 4 Weill Cornell Medical, 5 Department of Radiology, Dana-Farber Cancer Institute, Boston, MA, 6 Department of Radiology, Brigham and Women's Hospital, Boston, MA, 7 Harvard Medical School, Cambridge, MA

Objective

- The development of an NLP-based federated learning model for clinical risk prediction task is proposed as a solution to improve the accuracy and efficiency of clinical risk prediction.
- This model will be trained on locally stored clinical data at different healthcare institutions. Federated learning approach will be used for model training and testing to maintain patient privacy.
- By leveraging the power of NLP and the distributed nature of federated learning, this model has the potential to significantly enhance clinical risk prediction by leveraging diverse datasets while also preserving patient privacy.

Background

- Large language models have shown strong performance on various tasks including natural language understanding, natural language inference, named entity recognition and question answering (1-3). Pre-trained language models are used for the tasks as it helps in faster convergence.
- While their performance is typically strong, a major bottleneck for training large language models is the need for large datasets, which are expensive and time-consuming to gather and annotate. For certain data modalities, such as electronic health records and finance datasets, this problem is exacerbated due to restrictions and privacy concerns (4).
- We propose a federated training framework for language models, which uses pre-trained models for training on datasets held at each institution, which helps in data privacy and helps in distributed learning. This approach of distributed learning improves convergence speed and offers results comparable to centralized training.

Methods

- We use the Medical Information Mart for Intensive Care – III (mimic-III) dataset for predicting patient readmission. It consists of ~58 K electronic health records from ~38K patients in the intensive care unit between 2001 to 2012 (5). We used patient discharge notes for training. We split the discharge data into 3 subsets depending on the length of stay i.e., discharge notes at 3 days, 2 days and at end of stay.
- Pre-trained ClinicalBERT Model was used to predict patient readmission (6). The pretrained ClinicalBERT model weights encode information on clinical language. In order to adapt this model to our task, we perform model fine-tuning on the given task of patient readmission. The pretrained model requires less data for model convergence. Fine-tuning model weights on the similar task achieves better results than training a given model from scratch.
- The model takes patient clinical notes as input and outputs the patient's risk of readmission within 30-days. Model parameters are fine-tuned to maximize log-likelihood of this binary classifier. Final prediction for patient readmission was obtained by thresholding (> 0.5) output probability maps.
- Data is preprocessed as follows:
  a) words were converted to lowercase
  b) line breaks were removed
  c) special characters were removed
  d) sentence detection via SpaCy is used to segment notes
  e) rule-based modifications of special cases are made (e.g., M.D. replaced with MD, removing decimal-based numbers, replacing medication details)
- Rhino Health platform was used for all federated training and inference (7). Rhino Health uses Nvidia Flare platform for the federated distributed architecture. This helps in leveraging the benefits of Nvidia's architecture and Rhino Health's platform for an effective federated pipeline for the task of patient readmission (8-10).

Workflow

Figure 1. Flow-Chart of the Workflow for the Rhino Health Platform.

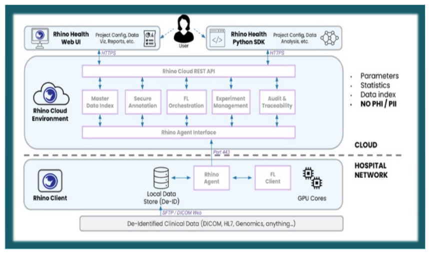


Figure 2. Flow-Chart of the NLP Trainer pipeline.

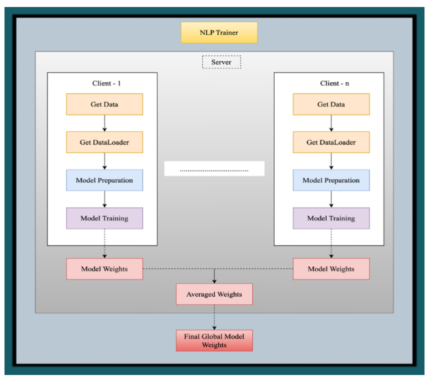


Figure 3. Flow-Chart of the NLP Inference pipeline.

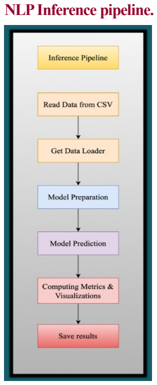


Table 1. Model Parameters used for Clinical-Bert Model

Model Parameters	Value
Hidden layer	12
Attention head	12
Hidden size	768
FFN* size	3072
Vocabulary size	30522
Max. position embeddings	512
Activation	gelu
Dropout	0.1
Total Parameters	**110M**

Data Distribution

- Mimic dataset containing the discharge summary was used for training purpose.
- Training dataset consists of 5006 samples, ~81 % of the dataset.
- Training data for the federated learning was randomly distributed between 2 client, and the data split was 50% between the clients.
- Validation set consists of 573 samples , ~9 % of the dataset.
- Testing set consists of 584 samples , ~9 % of the dataset.

Results

Table 2. Results obtained from centralized training framework vs federated training framework.

Metrics	Models	
	Centralized Model	Federated Model
Accuracy	65.75%	69.86%
F-1 Score	64.29%	69.76%
Sensitivity	62.07%	70.00%
Specificity	69.39%	70.00%
PPV	66.67%	69.52%
NPV	64.97%	70.21%
AUROC	0.71	0.75
AUPRC	0.66	0.72

- All results were obtained from the held-out testing set which was used to evaluate both training frameworks.
- AUROC value for the ClinicalBert model using centralized training was 0.71
- AUROC value for the ClinicalBert model using federated training with 2 clients was 0.75
- AUPRC value for the federated model was better compared to the centralized model by 0.6.
- F-1 score, accuracy, sensitivity and specificity for the centralized model was 64.29%, 65.75%, 62.07%, and 69.39% and for federated model was 69.76%, 69.86%, 70.00%, and 70.00% respectively.

Figure 4. Receiver Operating Characteristic (ROC) curve obtained using centralized training framework vs federated training framework.

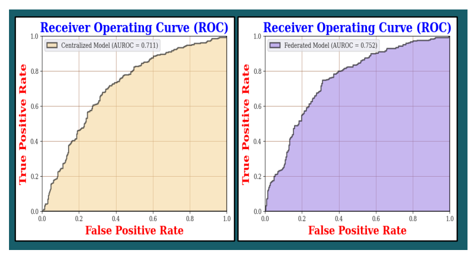
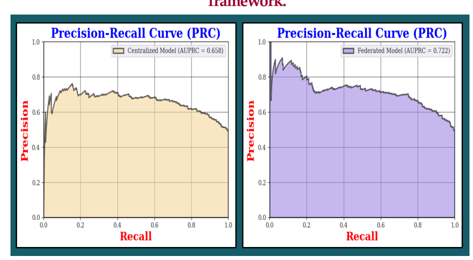


Figure 5. Precision Recall Curve (PRC) obtained using centralized training framework vs federated training framework.



Conclusions

- This work demonstrates the use of federated learning for training large language models on decentralized datasets across multiple hospitals or medical institutions.
- FL pipelines that leverage data from multiple hospitals will allow researchers to train more robust AI models that generalize to different patient demographics and have less bias.
- The advantage of our approach was use of limited resource (patient data and gpu's) for federated training and preserving data privacy for training and inference purpose, which is a key for medical institutions.
- All FL is done in a decentralized fashion, allowing hospitals to train stronger models while mitigating privacy concerns.

References

1. K. Clark et al., "Electra: Pre-training text encoders as discriminators rather than generators," arXiv preprint arXiv:2003.10555 (2020).
2. J. Devlin et al., "Bert: Pre-training of deep bidirectional transformers for language understanding," arXiv preprint arXiv:1810.04805 (2018).
3. Y. Tian et al., "FedBERT: when federated learning meets pre-training," ACM Transactions on Intelligent Systems and Technology (TIST) 13(4), 1-26 (2022).
4. A. Radford et al., "Language models are unsupervised multitask learners," OpenAI blog 1(8), 9 (2019).
5. A. E. Johnson et al., "MIMIC-III, a freely accessible critical care database," Scientific data 3(1), 1-9 (2016).
6. K. Huang, J. Altosaar, and R. Ranganath, "Clinicalbert: Modeling clinical notes and predicting hospital readmission," arXiv preprint arXiv:1904.05342 (2019).
7. K. Stephens, "Rhino Health Raises $5 Million to Improve AI Workflows in Healthcare Using Federated Learning," AXIS Imaging News (2021).
8. I. Dayan et al., "Federated learning for predicting clinical outcomes in patients with COVID-19," Nature medicine 27(10), 1735-1743 (2021).
9. H. R. Roth et al., "NVIDIA FLARE: Federated Learning from Simulation to Real-World," arXiv preprint arXiv:2210.13291 (2022).
10. D. Wu et al., "E-107 Multi-center study of a deep learning model for intracranial aneurysm detection in computed tomography angiography," British Medical Journal Publishing Group (2022).

Acknowledgments

- We would like to thank Alex Tonetti, M.S., Tal Einat, M.S., and Malhar Patel, M.D. from Rhino Health Team for their continuous support for this project.

Contacts

Sahil Nalawade
sahil_nalawade@dfci.harvard.edu
AI ML Scientist II
Dana-Farber Cancer Institute

Soujanya Samineni
Soujanya_Samineni@dfci.harvard.edu
Sr. Machine Learning Engineer
Dana-Farber Cancer Institute