

# Federated Learning for Privacy-Preserving AI: A Multi-Institutional Model for Automated Tympanic Membrane Perforation

Obinna I. Nwosu, MD<sup>1,2</sup>; Mitsuki Ota BS<sup>2</sup>; Lauren Buchholz BS<sup>3</sup>; Asitha D.L. Jayawardena, MD, MPH<sup>3,4</sup>; Aaron K. Remenscheider<sup>1,2,5</sup>; MD; Daniel J. Lee, MD<sup>1,2</sup>; Matthew G. Crowson, MD<sup>1,2</sup>.

## Background

- Gross visual estimation of tympanic membrane perforation (TMP) size is prone to error and can vary between practitioners.
- Developed a multi-institutional computer vision (CV) model for automated TMP size estimation.
- Utilized federated learning (FL) to train a global CV model while preserving patient privacy by keeping data decentralized.
- Instead of sharing data, FL workflows share model weights, or logic, learned from institutional training with the global CV model.
- This study demonstrates the feasibility of a privacy-preserving approach to training a global CV model for instance segmentation in otoscopic analysis.

## Methods

- Three CV models were developed independently using distinct datasets: public, institutions A, B, and C used 44, 57, 42, and 21 images each.
- TM and remnant tympanic membrane (RTM) boundaries were manually annotated on every image to create ground-truth annotations.
- Open source toolkit Detectron2 was utilized to train and test models.
- Utilized a standard FL workflow (Figure 1) with institutional weights used to update the global model each round
- After each round, global and institutional models were tested on the external/held-out validation dataset (Institution C).
- Accuracy and mean average precision (mAP-50) were used to evaluate model performance.

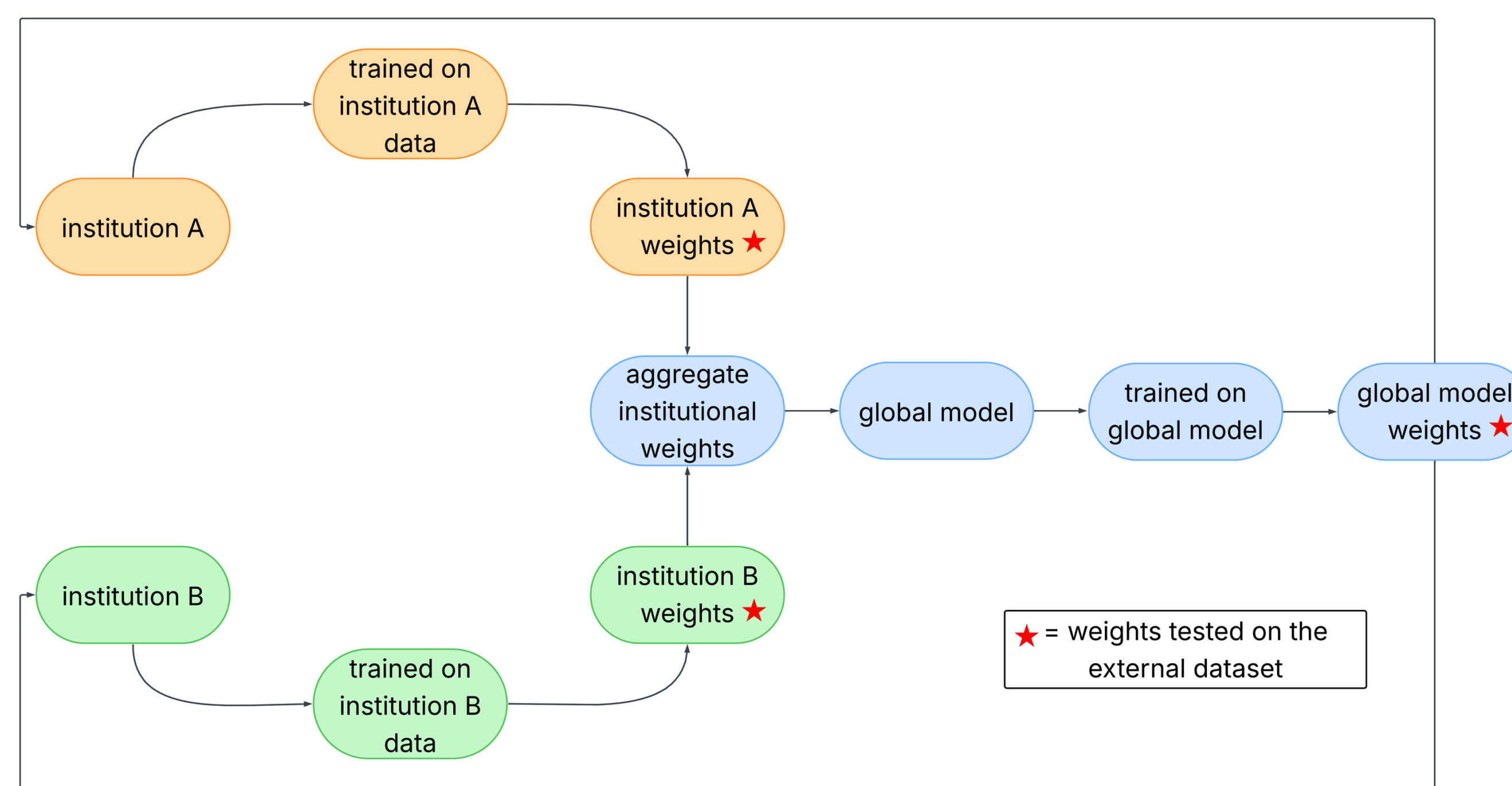


Figure 1. Flowchart depicting federated learning process. This was repeated 5 times.

### Definition of performance metrics:

- Accuracy:** the proportion of all predictions (TP + TN) that are correct.
- Precision:** the proportion of positive predictions that are correct (TP / (TP + FP)).
- Mean Average Precision (mAP50):** a measure that summarizes both precision and how well the model detects objects across different thresholds.

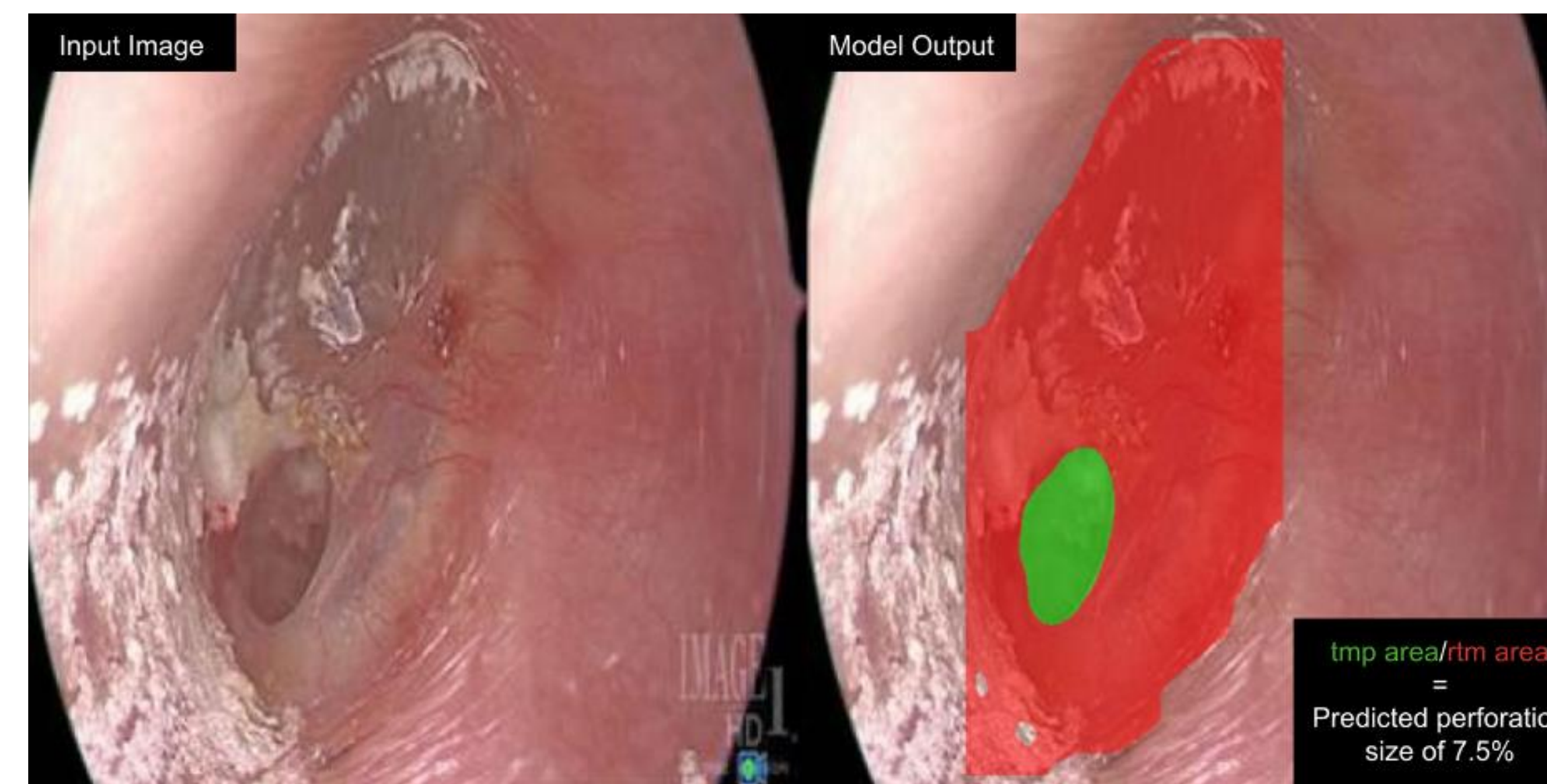


Figure 2. Example of global model performance on a held-out image, not encountered during training, from the external validation dataset.

## Results

- All models showed improved performance with successive FL rounds
- The global model achieved the highest mean average precision (mAP50) over 5 total FL rounds, though 95% confidence intervals overlapped between models.
- The final global model predicted TMP size in validation set with 92.5% accuracy (p value = 0.319)

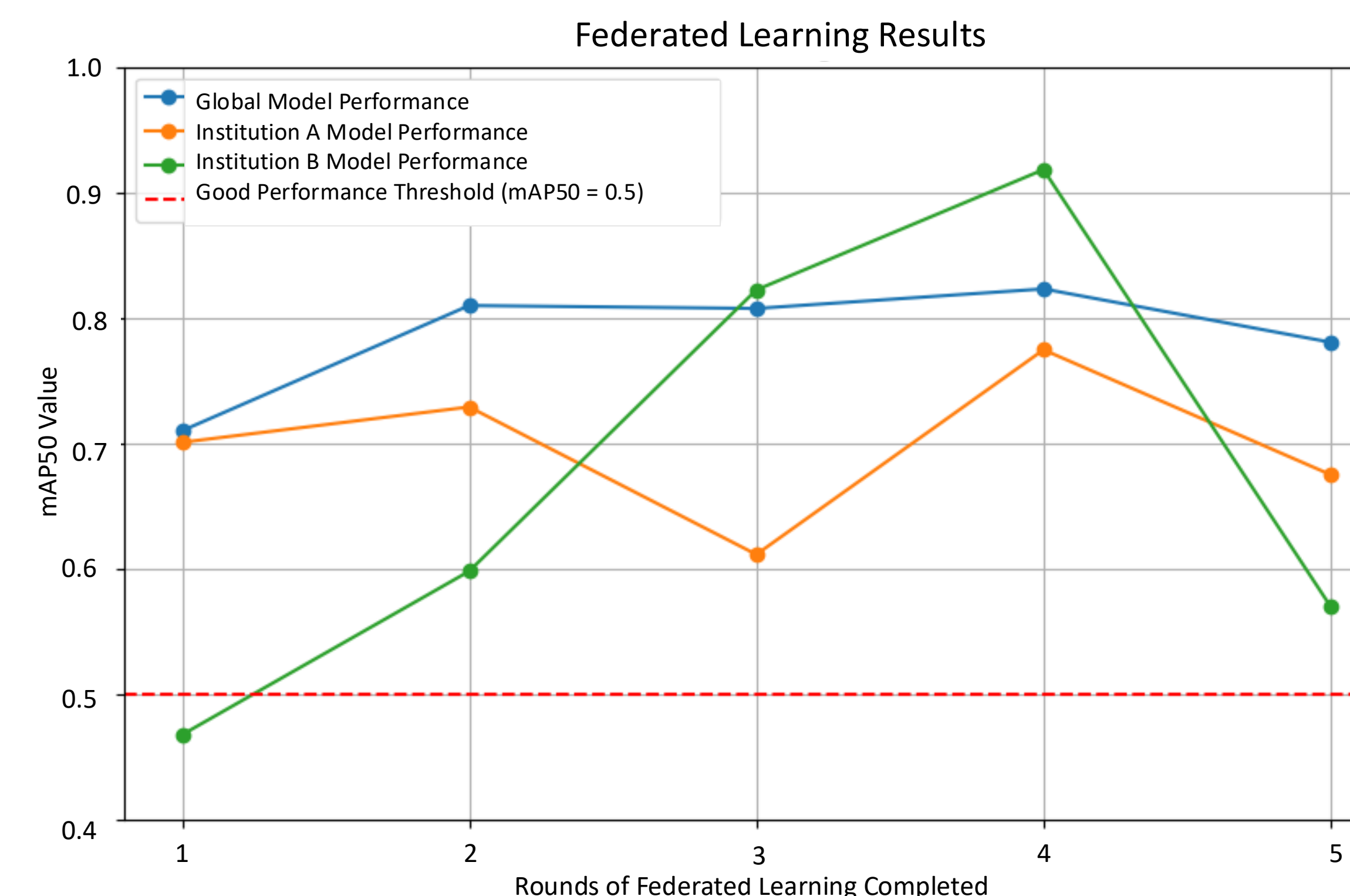


Figure 3. Trends of model performance over course of 5 rounds of federated learning.

## Discussion

- The global model outperformed all other models after 5 rounds.
- By sharing insights from local training, underperforming institution models can outperform other models.
- This study highlights feasibility of safe collaboration between independent institutions.
- Limitations:
  - Model generalizability is constrained by small training dataset (compared to 100k+ of images used for state-of-the-art CV models)
  - Repetitive training on same dataset (i.e., no new data introduced) → ↑ risk of overfitting.
- Future Directions:
  - Engage more institutions to introduce new insights, learned from novel data, into the FL workflow
  - Extend FL approach to enable development of other CV models for additional otolaryngologic predictive tasks.

## Conclusion

- FL enabled the creation of a global, cross-institutional CV model for TMP size estimation with use of small institutional datasets.
- In this approach, collaborative, privacy-preserving AI models can be developed by sharing only model weights rather than patient data.
- To validate clinical applicability, future model training should focus on regularly updating the global model with weights from a broader set of institutions.

## References

- Nwosu O, Suresh K, Knoll R, Lee DJ, Crowson MG. A proof-of-concept computer vision approach for measurement of tympanic membrane perforations. *Laryngoscope*. 2024;134(6):2906–2911. doi:10.1002/lary.31270
- Lerut B, Pfammatter A, Moons J, Linder T. Functional correlations of tympanic membrane perforation size. *Otol Neurotol*. 2012;33(3):379–386. doi:10.1097/MAO.0b013e318245cea5
- Sheller MJ, Edwards B, Reina GA, Martin J, Pati S, Kotrotsou A, et al. Federated learning in medicine: Facilitating multi-institutional collaborations without sharing patient data. *Sci Rep*. 2020;10:12598. doi:10.1038/s41598-020-69250-1
- Rieke N, Hancox J, Li W, Milletari F, Roth HR, Albarqouni S, et al. The future of digital health with federated learning. *NPJ Digit Med*. 2020;3:119. doi:10.1038/s41746-020-00323-1