Survey on Explainable Federated Learning

Nazreen Shah(PhD21122)
Sumedha (PhD21123)

INDRAPRASTHA INSTITUTE *of*
INFORMATION TECHNOLOGY **DELHI**

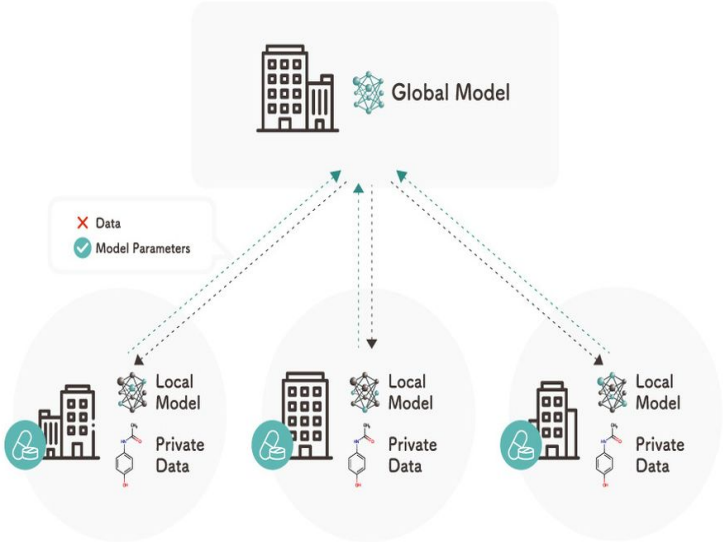

General
Data
Protection
Regulation

Outline

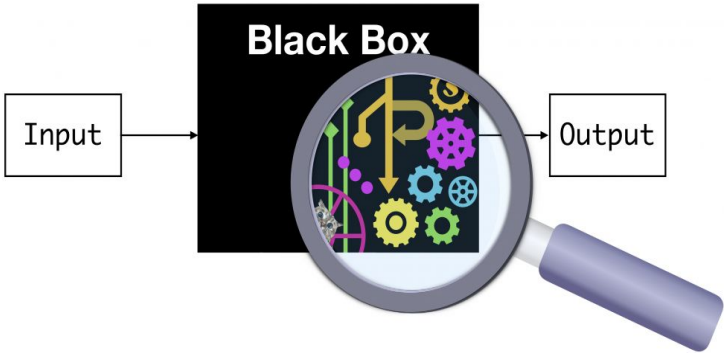
- Introduction
- Method
- Existing work
- Discussions
- Conclusion

Introduction

Federated Learning (FL) : Privacy preserving collaborative training of AI models.

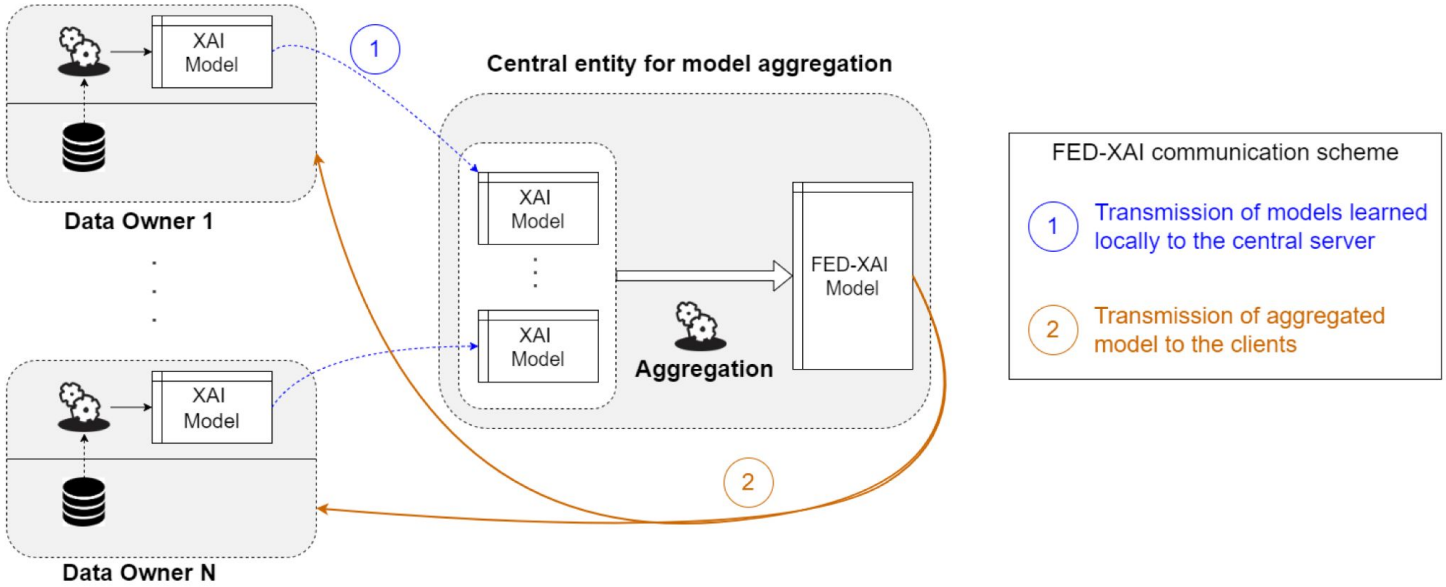


Explainable AI (XAI) : Enables humans to understand the outcome produced by Black Box AI models



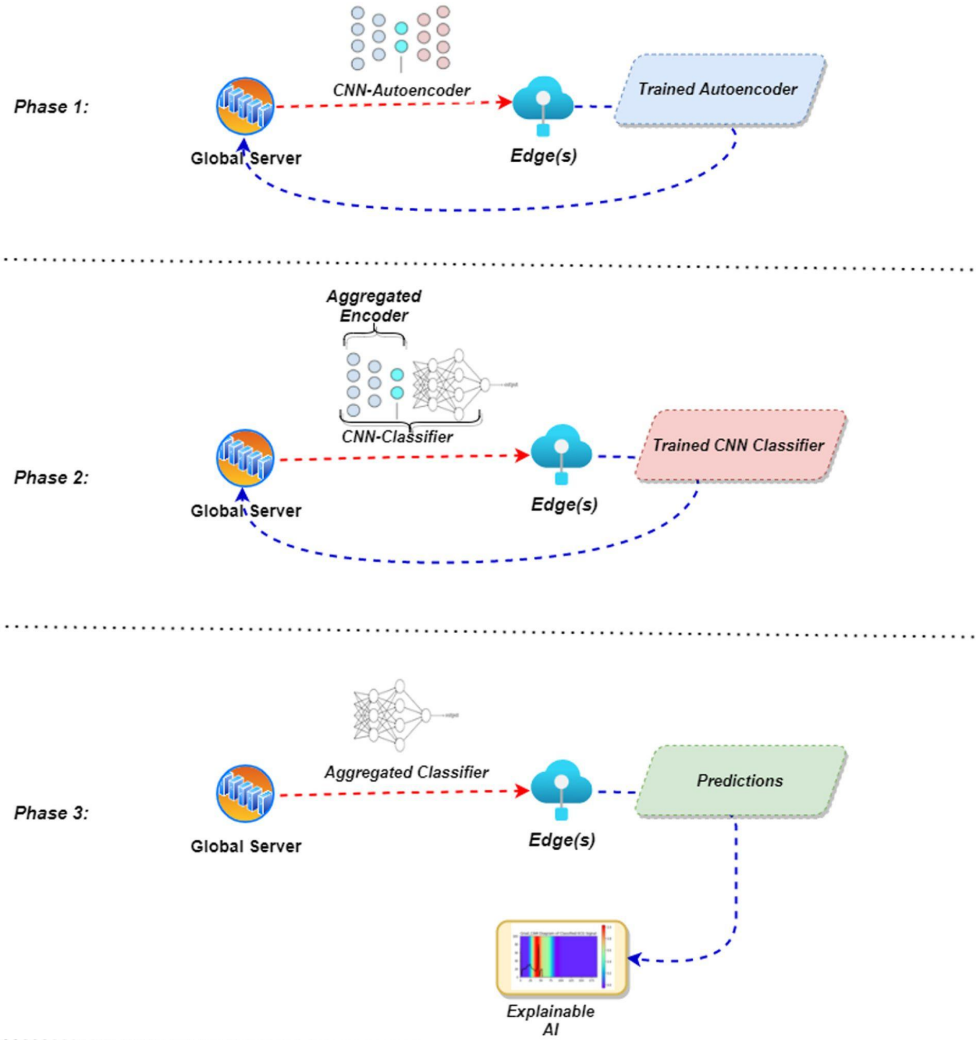
Methods

Inherently Explainable Methods



Methods

Post-hoc Methods



Existing Works

Work	Application	FL Feature	XAI Feature
[1]	Smart healthcare	For privacy and security	Inherently explainable
[2]	Regression problems	Vanilla FL	Fuzzy Rule-based Systems
[3]	6G-Automated vehicle networks	One shot communication FL	Inherently XAI
[4]	Data oriented AI systems	Vertical FL	Counterfactual explanation
[5]	Social media 3.0(IoT based)	Blockchain based, differentially private	Inherently explainable due to blockchains
[6]	Industry settings	Industry FL(IoT)	Dashboard visualization

Existing Works

Work	Application	FL Feature	XAI Feature
[7]	COVID-19 Detection	FedMoCo	GradCAM++
[8]	Personal Health Care	Horizontal FL	Feature relevance analysis and Visual explanation
[9]	ECG Based Healthcare	Federated Transfer Learning: FedMod	GradCAM
[10]	Taxi Travel Time Prediction	FedAverage	Integrated Gradients
[11]	Industrial Control System	FedeX	SHAP
[12]	Trustworthy FL	FederatedScope	Feature Importance MAP

Discussion

- There is a substantial lack of approaches for FL of inherently explainable models.
- Rule based methods cannot be integrated into FL using basic FL aggregation strategies due to its if-else modelling.
- The formulation of a differentiable global objective is impossible for these models
- We aim to propose a method with a twofold objective of privacy preservation (FL) and inherent explainability.

Conclusion

- Merging of XAI and FL is a big step towards data protection and explainability, leading us to Responsible AI.
- Thus, we studied & analyzed existing works in the area of ExplainableFL.
- We found a major gap in existing inherently explainable methods and their application in FL and as a future research direction, our goal is to tackle this challenge.

References

[1] A. Rahman et al., “Federated learning-based AI approaches in smart healthcare: concepts, taxonomies, challenges and open issues,” Cluster Computing, Aug. 2022, doi: https://doi.org/10.1007/s10586-022-03658-4.

[2] J. L. Corcuera Bárcena, P. Ducange, A. Ercolani, F. Marcelloni, and A. Renda, “An Approach to Federated Learning of Explainable Fuzzy Regression Models,” IEEE Xplore, Jul. 01, 2022.

[3] A. Renda et al., “Federated Learning of Explainable AI Models in 6G Systems: Towards Secure and Automated Vehicle Networking,” Information, vol. 13, no. 8, p. 395, Aug. 2022, doi: https://doi.org/10.3390/info13080395.

[4] P. Chen, X. Du, Z. Lu, J. Wu, and P. C. K. Hung, “EVFL: An explainable vertical federated learning for data-oriented Artificial Intelligence systems,” *Journal of Systems Architecture*, vol. 126, p. 102474, May 2022, doi: https://doi.org/10.1016/j.sysarc.2022.102474.

References

[5] S. Salim, B. Turnbull, and N. Moustafa, “A Blockchain-Enabled Explainable Federated Learning for Securing Internet-of-Things-Based Social Media 3.0 Networks,” *IEEE Transactions on Computational Social Systems*, pp. 1–17, 2021, doi: https://doi.org/10.1109/tcss.2021.3134463.

[6] M. Ungersböck, T. Hiessl, D. Schall, and F. Michahelles, “Explainable Federated Learning: A Lifecycle Dashboard for Industrial Settings,” *IEEE Pervasive Computing*, vol. 22, no. 1, pp. 19–28, Jan. 2023, doi: https://doi.org/10.1109/MPRV.2022.3229166.

[7] Nanqing Dong and Irina Voiculescu, “Federated Contrastive Learning for Decentralized Unlabeled Medical Images”, *MICCAI 2021: 24th International Conference*,pp. 378–387. Sep. 27–Oct.1, 2021, doi: https://doi.org/10.1007/978-3-030-87199-4_36

[8] A. Chaddad, Q. Z. Lu, J. L. Li, Y. Katib, R. Kateb, C. Tanougast, A. Bouridane, and A. Abdulkadir, “Explainable, domain-adaptive, and federated artificial intelligence in medicine,” IEEE/CAA J. Autom. Sinica, vol. 10, no. 4, pp. 859–876, Apr. 2023. doi: 10.1109/JAS.2023.123123

References

[9] Ali Raza, Kim Phuc Tran, Ludovic Koehl, Shujun Li, Designing ECG monitoring healthcare system with federated transfer learning and explainable AI, Knowledge-Based Systems, Volume 236, 2022, 107763, ISSN 0950-7051, https://doi.org/10.1016/j.knosys.2021.107763.

[10] Peng Chen, Xin Du, Zhihui Lu, Jie Wu, Patrick C.K. Hung, EVFL: An explainable vertical federated learning for data-oriented Artificial Intelligence systems, Journal of Systems Architecture, Volume 126, 2022, 102474, ISSN 1383-7621, https://doi.org/10.1016/j.sysarc.2022.102474.

[11] Fiosina, Jelena. “Explainable Federated Learning for Taxi Travel Time Prediction.” International Conference on Vehicle Technology and Intelligent Transport Systems (2021).

[12] T. T. Huong et al., "Federated Learning-Based Explainable Anomaly Detection for Industrial Control Systems," in IEEE Access, vol. 10, pp. 53854-53872, 2022, doi: 10.1109/ACCESS.2022.3173288.

Thank you !

Open to Feedback and Questions!

Nazreen Shah: nazreens@iiitd.ac.in
Sumedha: sumedhac@iiitd.ac.in