Title: Reviewing Federated learning and Explainable AI in medical imaging

Author:Addisu Amare

Date:03/06/2025

Abstract

Deep Learning Neural Networks are the primary framework used for analyzing medical images; however, the development of many new algorithms is limited by the use of data silos (data that is stored locally within institutions due to regulatory compliance) and the unique and opaque nature of most neural network models. In this article, we will discuss how Federated Learning decentralizes the training process, allowing several institutions to share their datasets safely. Additionally, we will discuss how Explainable AI enables the end-user to understand how AI makes its decisions, whether through techniques such as Grad-CAM or SHAP, and demonstrate how their combined power will provide transformative improvements, including 95-98% gains on benchmarks from tumor detection through chest X-rays[ieeexplore.ieee].

Introduction

Traditional AI methods rely heavily on a centralized system, which results in problems caused by fragmentation of medical data held by individual healthcare institutions. This type of centralization also poses a risk of data privacy violations under GDPR and HIPAA, and clinician acceptance is often Minimal due to the fact that understandability of results from "black-box" deep neural networks is limited. By utilizing federated learning (FL), organizations can collaborate to build AI models on real-world data without sharing the sensitive patient data itself. FL enables the training of AI to adequately address the challenges of non-independent and identically distributed data by taking advantage of the large amounts of diversity across multiple sites (centers) [pmc.ncbi.nlm.nih]. Additionally, explainable AI (XAI) has been developed to help solve the problem of model trustworthiness for physicians. XAI generates visual heat maps or saliency maps to visually demonstrate which pixels in an image contributed most to the prediction of an image. Ultimately, by addressing the existing long-standing issues of privacy, trustworthiness, and adoption, FL and XAI provide effective and interpretable diagnostic solutions for various medical conditions such as cancer and retinopathy. By working together, FL and XAI mark a new era in the delivery of ethical, scalable healthcare AI[mdpi].

Federated Learning in Medical Imaging

Personalised Federated Learning (FL) applies FL via localised training on siloed data to minimize communication while maintaining confidentiality. Applications include methods of brain tumour segmentation (e.g., Federated Averaging on the BraTS dataset) and COVID-19 detection from X-ray images. In both applications, these FL-based systems achieve a 5-15% improvement over comparable local models (AUC). Limitations, such as data heterogeneity, are overcome through personalised FL alternatives and quality weighting based on blockchain technology to provide broadly applicable models across different modalities [ieeexplore.ieee].

Challenge	FL Solution	Performance Impact [mdpi]
Data Privacy	Parameter-only sharing	HIPAA-compliant, no raw data transfer
Heterogeneity (non-IID)	Causal sparsification	+10% accuracy on multi-site CTs
Communication Overhead	Model compression	50-80% reduction in bandwidth
Low-Quality Sites	Blockchain valuation	Improved global F1-score by 7%

Explainable AI Techniques

Explainable AI (XAI) techniques such as LIME and Integrated Gradients demonstrate a model's weightings of feature values of images. This becomes increasingly important when validating both the pneumonia predictions of a model and the boundaries of lesions. Meanwhile, other architectures exist which inherentlyenable interpretability; they allow a heatmap to be closely aligned with the clinical observations found in oncology scans. For instance, Grad-CAM uses overlays to show MRIs where a radiologist would see tumour margins in 20-30% fewer cases of false positives[arxiv].

Integrating FL and XAI

The FedXAI framework uses global federated learning (FL) models to provide explanations after the fact, like creating a visual of how brain tumor classification works and providing a model with 97% prediction accuracy. Federated distillation allows for the retention of the interpretable quality during aggregation and Secure Multi-Party Computation protects user privacy on explanations. Through a successful combination of these frameworks, FedXAI effectively utilizes multi-modal evidence fusion—such as the use of PET-CT—to overcome challenges posed by evidence transferability; through evidential active learning[ieeexplore.ieee].

Framework	Key Innovation	Application & Gains [ieeexplore.ieee]
FedVGM	XAI-enhanced aggregation	Multi-dataset tumors: +5% mIoU
CFLM	Causal blockchain	Segmentation: 96% Dice score
FedLPPA	Weakly-supervised prompts	Reduced labels: 90% efficiency

Revolutionary Impacts

Not only do the two of these initiatives improve diagnostic capabilities by allowing for quicker and more accurate intervention through the federated XAI model's ability to identify retinopathy 25% faster than current methods across multiple clinics, but they also reduce the cost of annotation through the use of self-supervised FL. In addition, the deployment of ethical AI is increasing rapidly due to regulatory recognition of explainable federated tools in the EU and the USA[ijai.iaescore].

Challenges

XAI's computational overhead in real-time settings and FL's susceptibility to poisoning attacks (which is lessened by robust aggregation) are persistent problems. Multi-institutional benchmarks have standardization lags. [arxiv].

Conclusion

Federated learning and XAI, which combine privacy, power, and transparency to empower clinicians worldwide, usher in a reliable era for medical image analysis. Future paths indicate zero-trust architectures and multimodal, real-time systems that have the potential to save lives through democratized AI[frontiersin].