Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

Brain Tumor Detection using Federated Learning

CSE1908 - Capstone Project - Project Review 1

By:
Pratyush Kumar Singh (20MIA1131)

Guided By:
Dr. Amrit Pal

Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

INTRODUCTION

Federated learning (FL) is a cutting-edge machine learning paradigm that enables collaborative model training across multiple institutions without sharing sensitive data. This approach addresses significant privacy and governance challenges in healthcare by allowing local data to remain within institutional firewalls while still achieving high-performance models comparable to centralized data pooling methods.

In the context of any medical condition detection, FL offers promising applications in analyzing electronic health records, medical imaging, and signal data. It enhances the accuracy of disease diagnosis, supports personalized medicine, and facilitates real-time predictions. By leveraging FL, our project aims to develop robust models for detecting medical conditions while respecting patient privacy and data security.

LITERATURE REVIEW

Research Paper	Problem Addressed	Key Results	Dataset Used
Asymptotic Analysis of FL Under ETC	Communication-efficient FL with event-triggered SGD	Matches optimal convergence rates, flexible step sizes/triggers	
LIFL: Lightweight Serverless Platform for FL	Efficient serverless FL with hierarchical aggregation	2.7× faster, 5× less CPU than serverless; 1.6× faster than serverful	FEMNIST
Event-Driven Online VFL	Online VFL with event-driven activation and DLR	More stable, lower costs, sub-linear regret	i-MNIST, SUSY, HIGGS
Horizontal Federated Learning	Introduces HFL concepts and architectures	Summarizes privacy and efficiency advances from prior works	83% accuracy in classifying activities
FL: Strategies for Improving Comm. Efficiency	Communication-efficient FL techniques	Reduces communication by 100x with minor accuracy loss	CIFAR-10, Reddit Posts

Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

LITERATURE REVIEW

Research Paper	Problem Addressed	Key Results	Dataset Used
Federated Learning for Healthcare Domain - Pipeline, Applications and Challenges	How FL can be applied in healthcare to train models on distributed data while preserving privacy, and what challenges arise	FL effectively addresses privacy (e.g., GDPR, HIPAA compliance), outperforms centralized learning in privacy-sensitive cases; challenges include non-IID data, privacy risks, and computational costs	BraTS, EHRs, CIFAR-10
Federated Machine Learning in Healthcare: A Systematic Review on Clinical Applications and Technical Architecture	Investigating FL's clinical applications, technical robustness, and barriers to real-world healthcare adoption	Only 5.2% of 612 studies are real-life applications; FL is robust across data types (41.7% imaging) and models (76.3% neural networks); barriers include privacy breaches, infrastructure, and explainability	MRI/CT, tumor segmentation, COVID-19
Federated Learning: Overview, Strategies, Applications, Tools and Future Direction	Overview of FL's principles, strategies, and applications across domains (e.g., healthcare, IoT), with focus on privacy	FL strategies (e.g., FedProx, Scaffold) improve convergence on heterogeneous data; healthcare applications (e.g., EHR predictions) gain up to 10% accuracy; security risks persist but are mitigated by techniques	CIFAR-10(Strategy comparision), EHRs(Mortality, Disease prediction)
Vertical Federated Learning: Concepts, Advances	Comprehensive VFL review and VFLow framework	Synthesizes efficiency, effectiveness, privacy advances; highlights challenges	MIMIC-III, Avazu

Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

LITERATURE REVIEW

Research Paper	Problem Addressed	Proposed Solution	Dataset
Federated Learning in Health care Using Structured Medical Data	How FL can leverage structured medical data (e.g., EHRs) for multicenter clinical studies while preserving privacy and overcoming data-sharing challenges	Reviewed 23 studies; FL improves performance over local models (e.g., COVID-19 mortality prediction by Vaid et al. showed substantial gains); comparable to centralized models in some cases (e.g., Hansen et al.'s larynx cancer study); challenges include data heterogeneity and interpretability	MIMIC, Mount Sinai Health System EHRs, multi-institute EHRs
Federated Learning for Healthcare: A Comprehensive Review	How FL enables privacy-preserving deep learning in healthcare, addressing data security and collaboration across centers	Compared FL algorithms: FedAvg (82.74% server accuracy), FedPer (95.05% client accuracy), FedMA (77.91% server accuracy), Secret Sharing (faster, 85.35% server accuracy), Homomorphic Encryption (98% accurate, HIPAA/GDPR compliant); privacy preserved effectively	Human Activity Recognition dataset, TCGA
Federated learning-based AI approaches in smart healthcare: concepts, taxonomies, challenges and open issues	How FL and AI can enhance smart healthcare (e.g., IoMT, EHR management) while addressing privacy, security, and scalability issues	No experimental results; synthesizes FL-AI benefits: enhances privacy, reduces communication costs, balances accuracy-utility; identifies challenges like security, data heterogeneity; proposes future research directions (e.g., XAI integration)	IoMT, N-BaIoT
Privacy-Preserving Deep Learning: A Federated Learning Approach	Privacy issues in centralized deep learning models	FL framework combining differential privacy and secure multi-party computation	
Federated Learning for Healthcare: Systematic Review and Benchmarking	Challenges in deploying FL in healthcare	Systematic review and benchmarking of FL applications in healthcare	BraTS, COVID-19

SCOPE AND PROBLEM STATEMENT

The problem lies in the privacy concerns associated with centralized machine learning approaches for medical condition detection, which require aggregating sensitive patient data. To address this, we aim to develop a federated learning framework that enhances model accuracy while preserving patient privacy.

The scope of this project involves designing a federated learning system to train robust models for detecting specific medical conditions using decentralized data from multiple healthcare institutions. The focus will be on conditions like diabetes and cardiovascular diseases, using data types such as electronic health records and medical imaging. Key deliverables include a functional federated learning platform, performance evaluation, and best practices documentation.

Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

Dataset

- The dataset is a combination of the following three datasets :
  figshare
  SARTAJ dataset
  Br35H

- This dataset contains **7023** images of human brain MRI images.
- Classified into 4 classes:
  glioma
  meningioma
  pituitary
  no tumor

Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

RESEARCH CHALLENGES

- **Non-IID Data Handling:** Federated Learning (FL) systems often struggle with non-IID (independent and identically distributed) data, where the data distribution varies significantly across different clients. This can lead to difficulties in model convergence and reduced accuracy in detecting medical conditions effectively.
- **Resource Constraints on Edge Devices:** The implementation of FL on resource-constrained edge devices presents challenges in terms of computational power, memory, and energy consumption. Ensuring efficient processing while maintaining real-time performance is a significant challenge.
- **Model Personalization:** Achieving a balance between a generalized global model and client-specific personalized models is challenging. The heterogeneity in data across different environments requires models that are both accurate globally and effective locally.

RESEARCH OBJECTIVE

The primary objective of this research is to compare and evaluate the performance of machine learning models trained in traditional centralized environments with those trained using various federated learning (FL) algorithms for medical condition detection. Specifically, this study aims to assess the accuracy, efficiency, and privacy preservation of FL models in detecting specific medical conditions, such as diabetes or cardiovascular diseases, using diverse data types like electronic health records and medical imaging. The goal is to determine whether FL can achieve comparable or superior performance to traditional centralized approaches while maintaining enhanced data privacy and security. Additionally, the study will explore the challenges and opportunities associated with implementing FL in real-world healthcare settings.

LIMITATIONS AND FUTURE WORKS

- This project faces several limitations. The effectiveness of federated learning models depends on the quality and diversity of data, which can be challenging to obtain from multiple healthcare institutions. Privacy and security concerns remain, despite federated learning's benefits, and require robust measures to comply with regulations like HIPAA. Additionally, the computational resources needed for large-scale model training and coordination can be a challenge for institutions with limited infrastructure.

- Future research directions include expanding the federated learning framework to detect a broader range of medical conditions. Developing more robust privacy-preserving techniques, such as differential privacy, and improving model scalability will be crucial. Deploying the system in real-world healthcare settings and establishing multi-institutional collaborations can provide valuable insights and enhance the model's global applicability.

CONCLUSION

This project develops a Federated Learning-based framework in Healthcare, addressing key challenges of traditional disease detection systems like privacy concerns, data heterogeneity, and scalability. By enabling decentralized model training across various clients without sharing raw data, the framework enhances privacy and optimizes resource use. It effectively handles non-IID data and resource constraints. The project demonstrates the potential of Federated Learning to create secure, scalable, and privacy-preserving surveillance systems, offering a foundation for future advancements in diagnosis and effecient treatment technologies.

Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

GUIDE APPROVAL MAIL SNAPSHOT

REFERENCES

- Oh W, Nadkarni GN. Federated Learning in Health care Using Structured Medical Data. Adv Kidney Dis Health. 2023 Jan;30(1):4-16. doi: 10.1053/j.akdh.2022.11.007. PMID: 36723280; PMCID: PMC10208416.
- Rahman, A., Hossain, M.S., Muhammad, G. et al. Federated learning-based AI approaches in smart healthcare: concepts, taxonomies, challenges and open issues. Cluster Comput 26, 2271–2311 (2023). https://doi.org/10.1007/s10586-022-03658-4
- Dhade P, Shirke P. Federated Learning for Healthcare: A Comprehensive Review. *Engineering Proceedings*. 2023; 59(1):230. https://doi.org/10.3390/engproc2023059230
- Recent methodological advances in federated learning for healthcare Author links open overlay panelFan Zhang 1, Daniel Kreuter 1, Yichen Chen 1, Sören Dittmer 1 2, Samuel Tull 1, Tolou Shadbahr

Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

REFERENCES

- “Personalized Federated Learning with Exact Stochastic Gradient Descent“ Sotirios Nikoloutsopoulos, Iordanis Koutsopoulos, Michalis K. Titsias, arXiv:2202.09848 **[cs.LG]** (or arXiv:2202.09848v1 **[cs.LG]** for this version) https://doi.org/10.48550/arXiv.2202.09848
- “Federated Learning with Non-IID Data” Yue Zhao, Meng Li, Liangzhen Lai, Naveen Suda, Damon Civin, Vikas Chandra, arXiv:1806.00582 **[cs.LG]n**(or arXiv:1806.00582v2 **[cs.LG]** for this version), https://doi.org/10.48550/arXiv.1806.00582
- “A Performance Evaluation of Federated Learning Algorithms”, Adrian Nilsson, Simon Smith, Gregor Ulm, Emil Gustavsson, and Mats Jirstrand. 2018. A Performance Evaluation of Federated Learning Algorithms. In Second Workshop on Distributed Infrastructures for Deep Learning (DIDL ’18), December 10–11, 2018, Rennes, France. ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3286490.3286559

Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

REFERENCES

- G. Mathur and M. Bundele, "Research on Intelligent Video Surveillance techniques for suspicious activity detection critical review," 2016 International Conference on Recent Advances and Innovations in Engineering (ICRAIE), Jaipur, India, 2016, pp. 1-8, doi: 10.1109/ICRAIE.2016.7939467.
- **Agnostic Federated Learning**, Mehryar Mohri, Gary Sivek, Ananda Theertha Suresh arXiv:1902.00146 [cs.LG], (or arXiv:1902.00146v1 [cs.LG] for this version), , https://doi.org/10.48550/arXiv.1902.00146
- **Active Federated Learning:** Jack Goetz, Kshitiz Malik, Duc Bui, Seungwhan Moon, Honglei Liu, Anuj Kumar, arXiv:1909.12641 [cs.LG], (or arXiv:1909.12641v1 [cs.LG] for this version), https://doi.org/10.48550/arXiv.1909.12641
- Secure and Fast Asynchronous Vertical Federated Learning via Cascaded Hybrid Optimization arXiv:2306.16077 **[cs.LG]** (or arXiv:2306.16077v2 **[cs.LG]** for this version) https://doi.org/10.48550/arXiv.2306.1607

Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

REFERENCES

- S. Wang et al., "Adaptive Federated Learning in Resource Constrained Edge Computing Systems," in IEEE Journal on Selected Areas in Communications, vol. 37, no. 6, pp. 1205-1221, June 2019, doi: 10.1109/JSAC.2019.2904348. keywords: {Machine learning;Data models;Convergence;Distributed databases;Machine learning algorithms;Training;Peer-to-peer computing;Distributed machine learning;federated learning;mobile edge computing;wireless networking},"Federated Learning with Non-IID Data" Yue Zhao, Meng Li, Liangzhen Lai, Naveen Suda, Damon Civin, Vikas Chandra, arXiv:1806.00582 **[cs.LG]n**(or arXiv:1806.00582v2 **[cs.LG]** for this version), https://doi.org/10.48550/arXiv.1806.00582
- L. U. Khan et al., "Federated Learning for Edge Networks: Resource Optimization and Incentive Mechanism," in IEEE Communications Magazine, vol. 58, no. 10, pp. 88-93, October 2020, doi: 10.1109/MCOM.001.1900649. keywords: {Computational modeling;Collaborative work;Optimization;Data models;Servers;Internet of Things;Training},
- C. V. Amrutha, C. Jyotsna and J. Amudha, "Deep Learning Approach for Suspicious Activity Detection from Surveillance Video," 2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA), Bangalore, India, 2020, pp. 335-339, doi: 10.1109/ICIMIA48430.2020.9074920.

Thank you!