Research Article *Open Access*

Journal of Scientific Engineering Advances

ISSN: 3071-0723

DOI: *doi.org/10.63721/26JSEA0122*

Federated Deep Learning: Privacy Preservation in Multi-Site Soft Tissue Sarcoma Diagnosis

Majji Kavya, Donga Dhanushya, Pabbireddy Bhavya Sri Jyothi and Chandrasekhar koppireddy*

Department of Computer Science and Engineering Pragati Engineering College(A), Andhra Pradesh, India

Citation: Chandrasekhar koppireddy, Majji Kavya, Donga Dhanushya, Pabbireddy Bhavya Sri Jyothi (2026) Federated Deep Learning: Privacy Preservation in Multi-Site Soft Tissue Sarcoma Diagnosis. J. of Sci Eng Advances 2(1) 1-11. WMJ/JSEA-122

Abstract

STS are rare malignant disorders, and numerous histological subtypes exist that impede the diagnostic accuracy due to small datasets that are annotated by experts at institutions. The author presents a privacy-conserving federated learning architecture in this paper that enables the collaborative. training on a model across multiple medical institutions without breaching patient privacy. Our is an algorithm, which integrates the privacy of differentials, the homomorphic encryption, and the secure multi-party. computation. The experiments of multi-institutional histopathology data of large scale confirm. that federated learning can equally achieve diagnostics, as centralized models. It eliminates the privacy threat and, (87.3% vs. 89.1% accuracy) besides, it is more precise. The framework addresses non-IID. data distribution, communication efficiency, and resistance to attack vectors and demonstrates federal federated. learning to be a feasible solution to the diagnostics of rare cancer and satisfies the regulations at the same time.and governance enhancement.

***Corresponding author:** Chandrasekhar koppireddy, Department of Computer Science and Engineering Pragati Engineering College(A), Andhra Pradesh, India.

Submitted: 05.02.2026 **Accepted:** 11.02.2026 **Published:** 27.02.2026

Keywords: Deep Learning, Privacy Preservation, Federated Learning, Soft Tissue Sarcoma, Medical Image Analysis, Differentiation Privacy

Introduction

Background

Soft tissue sarcomas constitute approximately 1 per cent of malignancies in the adult population that have over 70 histological subtypes. Misdiagnosis is 25-40 at the communal hospitals. There is a prospect in deep learning with automated diagnosis, but there is a severe bottleneck in development: each year, individual. institutions receive insufficient STS cases and the traditional centralized machine.

Pooling of patient data would be required in learning, and this is a sensitive field of concern in terms of privacy and regulation in the HIPAA. GDPR, and similar frameworks. Solution Federated Learning.

Federated Learning Solution

Federated learning is the joint model training of distributed datasets without data communication. data sharing. Every location has models that are based on local training on proprietary datasets and only model. parameters are mixed to produce a global model that does not remain raw patient data. source institutions. This plan has the ability to draw in multi-institutional insight. datasets as well as maintaining privacy simultaneously. This contrast between centralized data pooling and decentralized collaborative training is illustrated in Figure. 1.

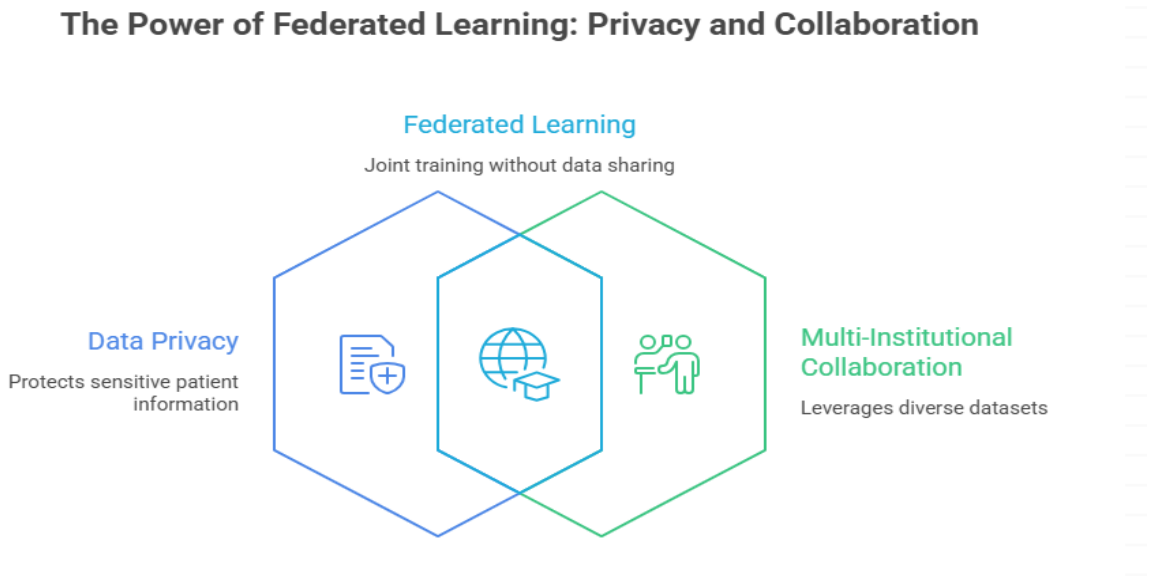

Figure 1: Centralized Versus Federated Learning Approach

Research Objectives

In this research, the researchers take critical needs into consideration since there are five objectives that the researchers have addressed:

- Development of federated learning systems of inter-site STS histopathological analysis,
- Including introduction of privacy. having mechanisms like differential privacy and secure aggregation,
- Exploring. plans of managing non-IID sites information,
- Empirical actual-data assessment, world datasets and (5) clinical analysis of privacy-utility trade-offs.

Related Work

Deep Learning for STS Diagnosis

The present paper presents the findings of the research that investigated the possibility of deep learning to diagnose STS. The recent researchers have shown that transfer learning is effective on models including ResNet, DenseNet, and EfficientNetEfficientNet STS classification, and their results achieve the accuracies of approximately 5-10 per cent. performance of an expert. pathologist. Most of the studies are however founded on uni-institution data that is insufficient. sample sizes. There is uncertainty as to whether the findings can be generalized to the external populations, and whether they will be affected by poor performance once introduced

to other institutions.

Federation Fundamentals of Learning

It is founded on Federated Averaging (Fed Avg) that is composed of repetitive sequence of local training. global aggregation, parameter transmission, and parameter transmission. These advanced algorithms are Fed Prox, Fed Nova, and SCAFFOLD that improve the convergence in heterogeneous settings. Issues such as non-Distribution of IID data, system and local data privacy weaknesses, and system heterogeneity are the key challenges.

Privacy Protecting Methods

Sensitive information can be leaked in a model parameter through membership inference. inversion attacks and property inference attacks. Privacy preserving methodologies will entail:

- Differential Privacy: The privacy is provided with the controlled noise that can be quantified. Passing privacy, utility, under epsilon and delta.
- Secure Multi-Party Computation: Cryptographic protocols which can be applied to do joint computation. and no disclosure of personal contributions.
- Homomorphic Encryption: Encrypts information, and the encrypted information can be computed without being decrypted. and despite introducing substantial computational overhead.

Healthcare Applications

Federated learning has been applied to the brain tumor segmentation, diabetic retinopathy, and COVID-19. It is recorded in research that there is performance which is equivalent to centralized forms, where heterogeneous distributions exist there are gaps. Research in request to rare diseases like STS. has not been adequately investigated even with the urgent cooperative learning demands.

Methodology

System Architecture

Client Architecture: In every institution with proprietary. training local models, computing updates to transmission and histopathological datasets, there exists a local node. It consists of data preprocessing pipelines, local CNN-based training modules. structures, gradient perturbation privacy, and communicating interfaces.
Server Architecture: Federation learning consists of a central server which does not access raw data. They include broadcasting of world parameters, gathering of encrypted/ privative updates. conducting of secure aggregation, and distributing updated models. The server is operating in the absence of individual update decryption.

Communication Protocol: The TLS is used due to the use of secure communications. Iterative procedure: by making use of initial construction, local training, transmission of updates, secure aggregation and global update. distribution.
The overall client–server interaction and secure aggregation workflow are shown in Figure. 2.

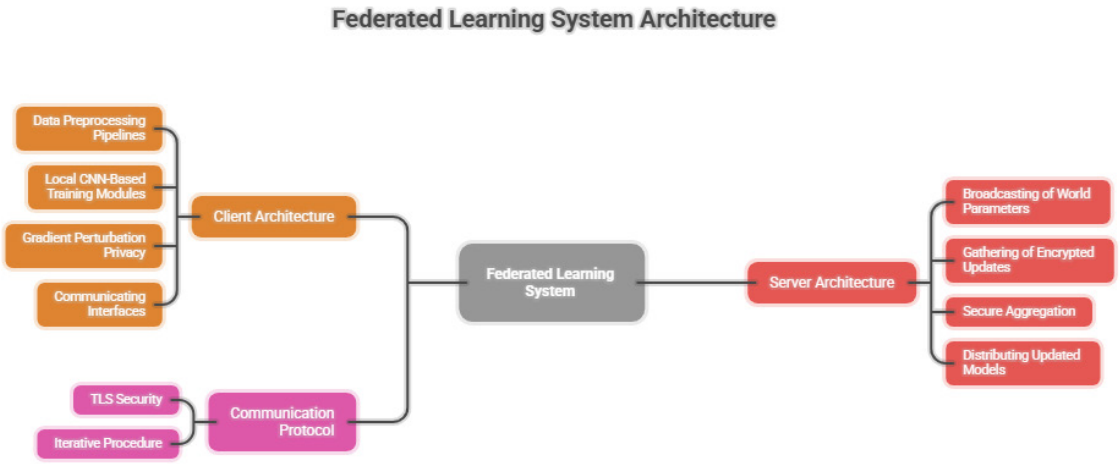

Figure 2: Federated Learning System Architecture

Deep Learning Architecture

Deep learning architecture is based on Adjusted ResNet-50 architecture that takes 224x224 RGB. histopathological patches as input. Architecture Architecture has skip connection of blocks, squeeze block, and-excitation block of feature enhancement and classification head of global average. pooling and dropout. The use of SoftMax in multi-class classification in major is applied to the output node. STS Liposarcoma, leiomyosarcoma, synovial sarcoma, malignant peripheral nerve. sheath tumor undifferentiated pleomorphic sarcoma.

Privacy Protection Operation

Differential Privacy: At the client side, it involves adding to the gradients before transmission, calibrated Gaussian noise N(0,0.5I) x, and 0.5 has been set to guarantee (epsilon, delta)-different privacy. The sensitivity is gradient clipped. The privacy budget was managed by the assistance of the Reyni Differential Privacy. of closer bounds of composition.

Secure Aggregation: secret sharing protocol makes a server incapable of viewing individual, gradients. Clients in the client model generate random masks randomly and shared in pairs, generate masked updates to the pairs and send them. masked gradients. Server, whose mask sums up to null.

Homomorphic Encryption: Pail Lier scheme encourages the capability to utilize encrypted aggregation of the gradient. maximum privacy guarantee. Clients, server performs the encryption of the public key updates. Non-decryption encryption operations.

Handling Heterogeneity of Data

Non-IID Data Problems: The institutional data are not homogenous (disease prevalence varies) are not homogeneous (features distribution varies) and are not homogeneous (pathologist expertise varies).

Mitigation Strategies:

- FedProx Algorithm: Adapts local objectives by including proximal term objectives keep away global model
- Adaptive Aggregation: Ranking clients by local validation performance.
- Data Augmentation: Stain normalization, domain shift Stain augmentation helps decrease domain shift.
- Individualised Federated Learning: Model and institution specific worldwide.

Communication Optimization

Communication Optimization, may be obtained through various techniques that include the use of different media, the use of software applications, conducting seminars and training the health caregivers.

Gradient Compression: Magnitude Sportifies the largest k% gradients. The process of quantization distorts the accuracy to either 8-bit or 4-bit values. A combination of both techniques saves bandwidth. Over 95 percent that have minor impacts.

Adaptive Communication: Involvement of clients with limited bandwidth is less or updates are compressed. Availability of updates is used to aggregate schedules in servers.

Experimental Setup

Datasets

Multi-Institutional Dataset: Pooled 15,847 histopathological images at 5 institutions. (A: 4,231 images, B: 3,892, C: 3,156, D: 2,734, E: 1,834). The five important STS subtypes projected. so that being projected in different prevalence sites constitutes realistic non-IID conditions.

Split of data: 70 percent of data will be in training, 15 in validation and 15 in testing of each of the individual institutions. Other held-out institution (Institution F, 1, 247 images) is a generalization test.

Preprocessing: The images were brought to 224224, and the stain was brought to normal.

Training configuration

Local Training: b5, batch size 32, SGD optimizer, learning rate 0.01, momentum 0.9.

Federated Learning: 100 rounds of communication with client response rate 100% (all institutions). participate in each round).

Parameters Privacy: Differential privacy = [1, 3, 5, 10, 1/infinity], 10 -5, gradient clipping threshold C = 1.0.

Baseline Comparisons

All the information is condensed in a single location (privacy upper bound, accuracy<|human|>Known as Centralized Learning. benchmark)

The training of local data is on independent local training within each institution.

- Federated Averaging: no privacy and Federated Avg.
- Federated Averaging: Federated FedAvg and privacy.
- FedAvg + Differential Privacy: FedAvg on varying privacy budgets.
- Federated (proximal regularised) algorithm.

Evaluation Metrics

Diagnostic Performance: Accuracy, precision, recall, F1-score, AUC-ROC, class wise, and macro-averaged.

Privacy Analysis: Threat to membership inference attack, gradient reconstruction. attacks. Measures of attack success.

Communicational Efficiency: This is a total number of bytes sent out, communications rounds off to convergence, bandwidth requirements.

Generalization: Held-out F performance of an Institution.

Results

Diagnostic Results

Primary Results

- Centralized learning: 89.1% accuracy (privacy risk baseline)
- The accuracy of federated learning (FedAvg): 87.3% (1.8% difference).

- Local averages of training: 76.4% accuracy (12.7% variation to centralized)
- FedProx: 87.8% (optimistic federated performance). As compared to local training, federated learning is much more effective and can be deemed akin to centralized performance, which is why the paradigm of collaborative diagnostics of rare cancers warrants the mentioned level of importance.

A comparative visualization of centralized, federated, and local learning performance is presented in Figure. 3.

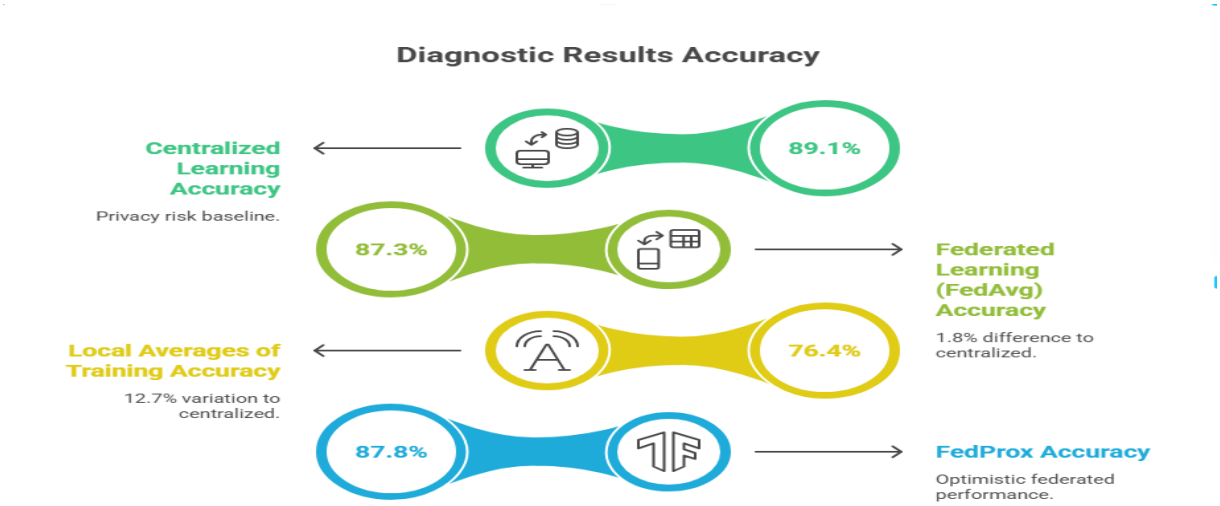

Figure 3: Diagnostic Performance Comparison of Different Learning Approaches

Per-Class Performance: Federated models can achieve the F1-scores: liposarcoma 0.89, leiomyosarcoma 0.86, synovial sarcoma 0.84, MPNST 0.82, undifferentiated pleomorphic. sarcoma 0.79. Less common subtypes represent bigger gaps than centralized learning due to small samples.

Privacy-Utility Trade-offs

Differential Effect in privacy:

- $\varepsilon = \infty$ (no privacy): 87.3% accuracy
- $\varepsilon = 10$: 86.8% accuracy (-0.5%)
- $\varepsilon = 5$: 85.5% accuracy (-1.8%)
- $\varepsilon = 3$: 84.1% accuracy (-3.2%)
- $\varepsilon = 1$: 79.7% accuracy (-7.6%)

The medium privacy budgets ($\varepsilon = 3$-5) are very protective and less accurate. degradation. Stricter privacy (=1) is highly detrimental to performance.

Attack Resistance: The membership inference attack success rates: the no privacy (5) 80, with 5 decreased to 54% (near random guessing 50%). Gradient reconstruction attacks were also effective with a gradient reconstruction attack, $\varepsilon \leq 5$.

Heterogeneity of Data Management

Non-IID Severity Analysis: There was also variation in the distribution of classes of institutions: Institution A<|human|>Institution A (specialized centre) had 45 percent cases of liposarcoma in comparison to Institution E (general hospital) which had 12 percent. Variations in characteristics between the various imaging protocols.

The performance of different federated learning algorithms under non-IID conditions is summarized in Table 1.

Table of Comparison of Algorithms under Non-IID:

- Fed Avg: 87.3% accuracy
- Fed Prox (0.01): 87.8% accuracy = 88.1% precision: Adaptive aggregation.

- FL with personalization: Accuracy (best) = 88.4%. The heterogeneity issues are conquered by means of improved techniques, in which case individualized techniques are applied. techniques which are most effective in regard to globalization and localization.

Algorithm Name	FedAvg	FedProx ($\mu = 0.01$)	Federated Learning with Personalization
Accuracy (%)	87.3%	87.8%	88.4%
Precision (%)	N/A	88.1%	N/A
Non-IID Data Handling Strategy	Basic global averaging	Adaptive aggregation using a proximal term to address client heterogeneity	Uses individualized local models to balance globalization and localization
Remarks	Limited performance under severe Non-IID conditions	Improves stability over FedAvg	Achieves the best performance under Non-IID data
Remarks	Limited performance under severe Non-IID conditions	Improves stability over FedAvg	Achieves the best performance under Non-IID data

Table 1: Comparative Analysis of Federated Learning Techniques

Communication Efficiency

Compression Results

- Uncompressed: 98.3 MB per round
- Top-10% sparsification: 9.8 MB (-90%)
- 8-bit quantization: 24.6 MB (-75%) Combined (sparse + quantized): 2.5 MB (-97.5%)

This compression cost has lost 0.8% accuracy (86.5%), and takes 97.5% bandwidth reduction (versus 87.3%), and thus can be used in low-bandwidth environments. Convergence Analysis: Fed Avg approaches to the normal distribution within 100 round of convergence. Communication-efficient compression variant converges in 120 rounds (20 percent reduction) but bandwidth is reduced. by 95 percent+ because of per-round savings.

External Site Generalization

Held-outs in Performance of the Institution:

- Centralized model: 86.3% accuracy
- Federated model 84.7% accuracy (1.6% gap)
- Institution F local model: 79.2% accuracy (7.1% gap vs. federated)

The comparative generalization performance across models is illustrated in Figure. 4. Federated models exhibit huge tendencies in their generalization to non-trained institutions, much higher external data training than local training. In various healthcare settings, very important in clinical practice. It will require calculation of the required parameters using a calculator and a spreadsheet.

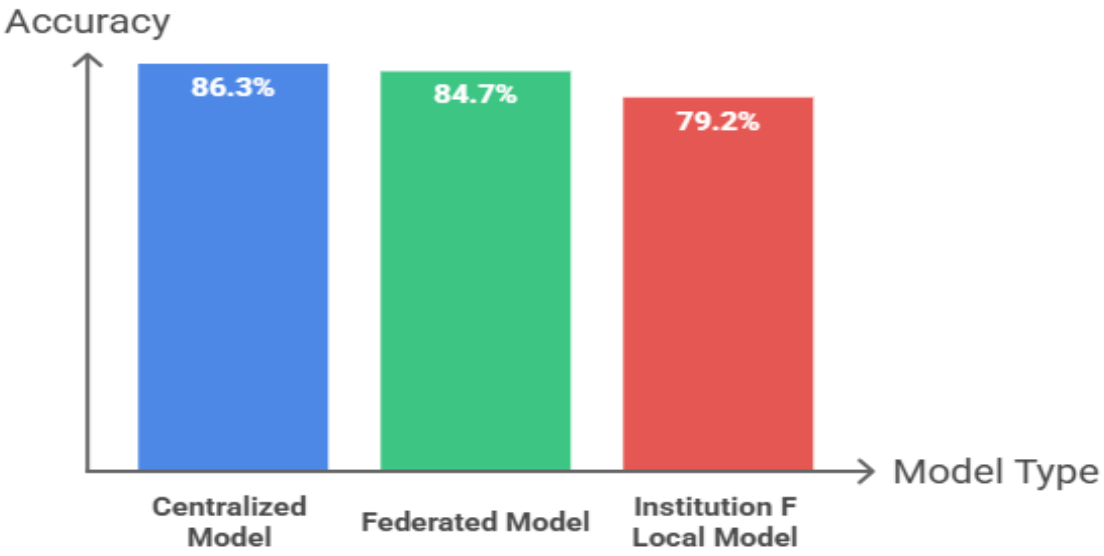

Figure 4: Model accuracy Performance Comparision

Computational Requirements

Training Time: Round average hours in training locally: 45 minutes across the institutions (variable depending on). on the local data volume and hardware). Total fed training: 100 round parallel training of 75 hours. Local training.

Resource Requirements: GPU memory: 8-12 GB. CPU cores: 8-16. Storage: 50-100 GB per institution. Achievable with standard devices of deep learning.

Discussion

Key Findings

Major Results Federated learning achieves an equivalent accuracy to centralized training on diagnostic tasks. and at the cost of annihilating privacy. Moderate budgets create high differentials on privacy. security having a limited consequence of accuracy. The model is information strong. adaptive aggregation and personalized learning. Optimization of communication enables it to be used in a bandwidth limited environment. On high secondary generalization to external. institutions checks applicability of the clinical.

Privacy-Utility Trade-offs

At 4-6 range: strong privacy protection strong privacy protection (membership inference close to): This range represents a hard privacy assurance. Stricter privacy (1) results in large degradation but looser budgets (10) result in insignificant additional privacy benefit. The guidelines used in selection of privacy parameters are dependent on application sensitivity, regulatory. requirements and acceptable accuracy.

Clinical Implications

Federated learning enables collaborative diagnostics of rare diseases in an institutional data. scarcity barrier. Generalizations are made when compared to training in one institution and this is crucial in the deployment in a variety of clinical settings. The privacy protection deals with the patient. issues and regulatory compliance. This entails institutional facilities, technical proficiency and execution. governmental frameworks.

Restrictions

Limitations of Data: The authors focused the study on five institutes and five major STS subtypes. Even the ultra-rare subtypes with less than 100 cases in the world are hard to manage. Simulation experiments needed. increased coverage of subtypes.

Computational Barriers: The privacy mechanisms incur 15-30 percent overheads. Few resource institutions may have a problem with participation. Democratization of access Cloud-based federated Learning platforms could be used to do the democratization of access.

Technical Problems: Assumes that there is a good network connection and synchronized participation. The applications in the real-life must address disconnected operation, asynchronous data. dropout.

Assumptions Privacy: Curiosity but honesty. Complex attacks (e.g. It needs more protection against Byzantine attacks, backdoor poisoning) etc. Privacy assurances require appropriate selection of parameter and proper implementation.

Future Directions

More Sophisticated Privacy Techniques: Privacy amplification by local differential privacy, Local differential privacy. randomisation and better composition theorems have the potential to improve protection or usefulness. Verification of privacy-preserving models will enable the participants to verify the correctness of the global model.

Scalability: Learn on larger networks of institutions in a hierarchical manner. Co-ordination via blockchain decentralization. Federation of intercontinental learning requires the best protocols.

Clinical Validation: Prospective study based on diagnostic workflow, clinical evaluation of reality. impact of decision-making and patient outcomes to be implemented.

Conclusion

The research confirms that it is possible to achieve federated deep learning privacy preserving in multi. diagnosis of institutional soft tissue sarcoma. The federated learning pre-pilot clinical preparedness is certified by achieved diagnostic performance (87.3 percent vs. 89.1percent), privacy protection (membership inference is close to random), communication, and centralized certify. deployments.

The comprehensive privacy-preserving system that is custom-crafted to identify rare cancers, critical empirical analysis of privacy utility trade-offs, new. methodologies of managing a non-IID medical information, and guidelines to be applied in practice towards privacy parameters. selection in clinical practice, are some of the major contributions. Where structural limitations to the development of AI in rare diseases is the absence of such data, federated learning makes it possible to achieve this in a mutual manner without violating the patient and privacy.

By engaging with the global institutions and establishing new principles of cooperation, this paradigm will be able to transform the research on rare diseases, improve health equity throughout the world. Artificial Intelligence in healthcare and privacy and ethics. Successful research-to-practice translation requires collaboration over the long term between machine. learning researchers, medical practitioners, privacy researchers, regulations, and patients. Such technical advancement must be accompanied with corresponding governance, rules and regulations, etc. clinical validation to enable federated learning to fulfil its promise of improved patient care in the presence of. protecting privacy value in the age of precision medicine.

References

1. Maussion C, Coindre JM, Blay JY (2025) Multimodal prediction of metastatic relapse using federated deep learning in soft-tissue sarcoma with a complex genomic profile. Scientific Reports 15.
2. Shukla S, Rajkumar S, Sinha A (2025) Federated learning with differential privacy for breast cancer diagnosis enabling secure data sharing and model integrity. Scientific Reports 15.
3. Yahiaoui M E, Derdour M, Abdulghafor R (2024) Federated Learning with Privacy Preserving for Multi-Institutional Three-Dimensional Brain Tumor Segmentation. Diagnostics 14.
4. Hu J, Yang Z, Wang P (2025) Federated Learning for Medical Image Analysis: Privacy-Preserving Paradigms and Clinical Challenges. Transactions on Artificial Intelligence 1: 153-169.
5. Koutsoubis N, Waqas A, Yilmaz Y (2025) Privacy-preserving Federated Learning and Uncertainty Quantification in Medical Imaging. Radiology: Artificial Intelligence 7: e240637.
6. Sajid Nazir, Mohammad Kaleem (2025) Federated Learning for Medical Image Analysis with Deep Neural Networks – a review of federated

learning and privacy preservation challenges. Diagnostics 13
7. Voigtländer H, Kauczor HU, Sedaghat S (2025) Diagnostic utility of MRI-based convolutional neural networks in soft tissue sarcomas: a mini-review. Frontiers in Oncology 15.
8. Rehman M H U (2023) Federated learning for medical imaging radiology: current developments, limitations, and future directions. British Journal of Radiology 96.
9. Ng D, Taylor M, Xiang Lan, Melissa Min Szu Yao, Mengling Feng (2025) Federated Learning: a collaborative effort to achieve better medical imaging models for distributed sites with small labelled datasets. Quantitative Imaging in Medicine and Surgery 11.
10. Liudajun, Han Ying, Hu Dongyan, Pan Fei (2025) Federated learning and differential privacy for medical image analysis. Digit Health 11.
11. Daniel Kwame, Aisha Bello, Jin-soo Park, Omar Al-Farouq (2026) Federated Learning in Medical Imaging for Privacy-Preserving Diagnosis. ResearchGate Review Article.
12. Nikolas Koutsoubis, Ghulam Rasool, Matthew Schabath, Ravi P Ramachandran, Yasin Yilmaz (2024) Future-Proofing Medical Imaging with Privacy-Preserving Federated Learning and Uncertainty Quantification: A Review. Electrical Engineering and Systems Science.
13. Zhou L, Wang M, Zhou N (2024) Distributed Federated Learning-Based Deep Learning Model for Privacy MRI Brain Tumor Detection. Electrical Engineering and Systems Science. https://arxiv.org/pdf/2404.10026.
14. Wakili A A, Adamu Hussaini, Abubakar A Musa (2025) TwinSegNet: A digital twin-enabled federated learning framework for brain tumor analysis. Electrical Engineering and Systems Science. http://arxiv.org/pdf/2512.17488.
15. Shiranthika C, Zahra Hafezi Kafshgari, Hadi Hadizadeh, Parvaneh Saeedi (2025) MedSegNet10: A repository for split federated medical image segmentation. Electrical Engineering and Systems Science 1-20.
16. Harikrishna Bommala, Sirisha Yerraboina (2024) Soft tissue sarcoma diagnosis using combined deep learning and clinical features. MATEC Web of Conferences 392.
17. Xie H (2024) Deep learning driven diagnosis of malignant soft tissue tumours using US images. Front Oncol 14.
18. Xu R, Tang J, Li C (2024) Deep learning-based AI for assisting diagnosis and treatment in soft tissue sarcomas. Meta-Radiology 2: 100-069.
19. Naz N S, Mehmood M H, Ahmed F (2025) Privacy preserving skin cancer diagnosis through federated deep learning and explainable AI. Scientific Reports 15.
20. Fan S, Awais Ahmed, Xiaoyang Zeng, Rui Xi (2025) A personalized multimodal federated learning framework with cross-modality knowledge transfer. MDPI Electronics 14.