QUBAHAN TECHNO JOURNAL
VOL. 2, Nov 2025
https://doi.org/10.48161/qjms.v2a85

PRIVACY-PRESERVING FEDERATED DEEP LEARNING FOR ROBUST MEDICAL IMAGE SEGMENTATION ACROSS MULTI-INSTITUTIONAL CLINICAL DATASETS

Ahmad Pitra Jumli[*1]

[1] School of Hospitality & Creative Arts, Management and Science University, Shah Alam, Malaysia;

* **Corresponding author:** gulmmira@yandex.ru.

ABSTRACT: Early diagnosis and therapy planning in brain tumor treatment depend heavily on medical image segmentation. However, Traditional DNNs require data from medical practices to be collected in a central location, raising significant privacy concerns about patient data leakage, security risks, and institutional data-sharing limits. This study proposes a robust, multi-institutional medical image segmentation system that uses federated deep learning while maintaining the privacy of clinical data. The proposed system combines a U-Net-based convolutional neural network with a federated learning architecture. The goal is to obtain multi-institution segmentation models without directly sharing healthcare institutions' original MRI data. To enhance data protection, differential privacy measures, such as gradient clipping and the injection of Gaussian noise, were applied during model optimization and the exchange of model parameters. The brain tumor MRI dataset from BraTS 2020 was used for experiments and distributed across four simulated healthcare clients to mimic diverse institutional environments. The experimental implementation was carried out using Python and PyTorch on Google Colaboratory. The segmentation results were evaluated with Dice Similarity Coefficient (DSC) and Intersection over Union (IoU). The Dice and IoU values were 0.3628 and 0.2464 for the centralized U-Net, and 0.3295 and 0.2176 for the federated U-Net, respectively, indicating that the federated U-Net achieved the highest segmentation performance while maintaining effective collaborative learning and privacy preservation. While the DP-Federated model offered stronger privacy protection, segmentation accuracy was compromised, with lower Dice and IoU scores of 0.0408 and 0.0211, respectively, due to noise perturbation. The results suggest a federated learning model with suitable segmentation capabilities and a significant reduction in exposure to sensitive medical data.

Keywords: privacy-preserving AI, federated learning, medical image segmentation, differential privacy, U-Net architecture, brain tumor MRI.

I. INTRODUCTION

1. BACKGROUND OF MEDICAL IMAGE SEGMENTATION

Medical image segmentation is a fundamental problem in medical image analysis, in which medical images are divided into meaningful regions such as organs, tissues, tumors, and lesions. It is essential for disease diagnosis and treatment planning and is used during surgery and clinical decision-making. The numerous medical imaging technologies, such as Magnetic Resonance Imaging (MRI), Computed Tomography (CT), X-ray, and Ultrasound, produce very large amounts of medical data that are critical for making accurate diagnoses and patient care decisions (Hussain et al., 2022). According to Veiga-Canuto et al. (2022), manual segmentation by radiologists is time-consuming and labor-intensive, and inter-observer variability exists, necessitating automated segmentation techniques.

With recent advances in Artificial Intelligence and Deep Learning, the performance of segmentation in medical images has been significantly enhanced. Medical image analysis with CNNs, including models such as Attention U-Net and U-Net, has shown great success in detecting intricate anatomical structures (Azad et al., 2024). Recent developments in segmentation have included the introduction of Transformer-based models, which have proven effective at capturing long-range dependencies between image elements. These developments are still insufficient to ensure robust, widespread segmentation across datasets from different healthcare institutions, which reflect differences in imaging protocols, equipment, and patient demographics. Thus, it is essential to develop intelligent, scalable segmentation systems, especially for modern healthcare applications.

2. IMPORTANCE OF FEDERATED LEARNING IN HEALTHCARE

In the medical field, Federated Learning (FL) has become a significant method in artificial intelligence due to its ability to train models collectively without sharing confidential patient information (Gu et al., 2023). In the traditional centralized learning approach, all medical data must be sent to a central server. Thus, it poses patient privacy and data security issues as well as regulatory compliance concerns. Healthcare organizations may have to comply with regulations like HIPAA and GDPR that can impose limitations on sharing clinical data with one another. Local models can be trained using Federated Learning by communicating only the model parameters or parameter updates to a central server, using local data at each hospital. This decentralized learning strategy will ensure data confidentiality and minimize privacy risks (Myakala et al., 2024). In medical image analysis, federated learning can facilitate the collaborative training of powerful, generalizable segmentation models using medical datasets distributed across multiple institutions. Lastly, federated learning is a promising approach that offers a secure, efficient, and scalable way to collaborate across healthcare organizations, enhance model accuracy, and foster the integration of AI systems in clinical settings while maintaining privacy.

3. PROBLEM STATEMENT

For deep learning-based medical image segmentation models to achieve high accuracy and robust generalization, large and diverse datasets are needed. However, medical data is often stored across various hospitals and healthcare institutions, where legal and privacy laws prohibit direct data sharing. The centralized learning model involves sending sensitive medical data to a central server, which poses serious privacy and security risks and may result in non-compliance with healthcare regulations. These difficulties hinder the development of successful, collaborative AY systems in clinical settings. Furthermore, diagnostic information from diverse institutions is frequently acquired using different imaging devices, acquisition protocols, annotation standards, and patient populations.

Non-identically distributed (non-IID) data is a consequence of this heterogeneity and negatively affects the performance and generalization of segmentation models (Zhang et al., 2025). Many existing segmentation systems lack robustness and cannot maintain patient confidentiality across a variety of clinical datasets. While federated learning can provide a decentralized solution for collaborative training, existing federated segmentation methods still face issues such as communication overhead, unstable model convergence, privacy exposure, and performance degradation in heterogeneous data settings (Fareed et al., 2025). Accordingly, there is a need for a federated deep learning framework that can enable secure, precise, and powerful medical image segmentation across multi-institutional clinical datasets while maintaining privacy.

4. AIM AND OBJECTIVES

The study aims to develop a privacy-preserving federated deep learning framework for robust and accurate medical image segmentation across multi-institutional clinical datasets without centralized data sharing.

5. RESEARCH OBJECTIVES

- To design a federated deep learning architecture for collaborative medical image segmentation using distributed clinical datasets.
- To integrate privacy-preserving mechanisms such as differential privacy and secure aggregation into the federated framework.

- To evaluate the segmentation performance of the proposed framework on MRI, CT, and ultrasound datasets collected from multiple institutions.
- To compare the proposed federated model with centralized deep learning approaches in terms of accuracy, robustness, scalability, and privacy preservation.
- 1.6 Research questions
- How can federated deep learning improve medical image segmentation while preserving patient privacy?
- Which privacy-preserving techniques are most effective for secure collaborative medical AI training?
- How effectively does the proposed framework perform across heterogeneous multi-institutional imaging datasets?
- How does the proposed federated model compare with centralized segmentation approaches in terms of segmentation accuracy and privacy protection?

7. SIGNIFICANCE OF THE STUDY

The study is significant because it addresses one of the most pressing issues in healthcare AI: creating precise medical image segmentation systems without compromising patient privacy. This future-oriented privacy-preserving federated deep learning approach has the potential to facilitate the secure and ethical deployment of AI in healthcare. The study helps reduce problems with inter-institutional collaboration, model generalization, and diagnostic reliability across different medical datasets. Technologically, the research combines cutting-edge deep learning segmentation models with federated learning and privacy-preserving technologies such as Differential Privacy (DP) and Secure Aggregation (SecAgg). This combination boosts the model's performance and secures data. The study also helps develop more robust AI systems that adhere to healthcare standards and support secure, large-scale healthcare data use.

II. LITERATURE REVIEW

1. DEEP LEARNING IN MEDICAL IMAGING

In the medical imaging field, deep learning has revolutionized the analysis, feature extraction, and disease detection in complex medical images through automated processes. Handcrafted features and manual interpretation were the traditional approaches to image processing; however, with the development of artificial intelligence, Convolutional Neural Networks (CNNs) have shown remarkable performance in medical image classification, detection, segmentation, and diagnosis. These models feature automatic learning from imaging data, yielding hierarchy-based features that improve accuracy and reduce manual effort (Salman & Kalakech, 2024).

However, deep learning applications are also prevalent across multiple modalities, including magnetic resonance imaging (MRI), computed tomography (CT), X-ray, Ultrasound, and histopathological imaging. It is used in a variety of applications such as tumor detection, organ segmentation, lesion identification, and disease progression analysis (Hussain et al., 2024). Medical image analysis tasks have been successfully tackled using architectures such as U-Net, ResNet, and DenseNet due to their feature-learning capabilities (Sistaninejhad et al., 2023). More recently, Transformer-based architectures have gained ground as well because they can represent long-range spatial context in images. However, these progressions still have drawbacks, such as insufficient data labeling, computational complexity, limited data diversity, and limited model transferability across different clinical settings.

2. CNN AND TRANSFORMER-BASED SEGMENTATION METHODS

In recent years, modeling the spatial and hierarchical features of medical images has gained significant attention, enabled by a new family of deep learning networks called Convolutional Neural Networks (CNNs). CNN-based models such as U-Net, UNet++, and Attention U-Net are widely used for tumor, organ, lesion, and tissue segmentation in MRI, CT, and X-ray images (Abueed et al., 2025). U-Net presented an encoder-decoder structure with skip connections, enabling its use in biomedical image segmentation to achieve high localization and boundary preservation. The attention-based CNN models were also tested to improve segmentation performance by focusing on clinically relevant regions and ignoring irrelevant background information. Although they work well, CNN models primarily focus on local spatial patterns and may fail to capture long-range contextual dependencies in complex medical images.

These limitations have been overcome by transformer-based architectures that incorporate self-attention mechanisms to capture global image context.

Models such as Vision Transformer (ViT), TransUNet, Swin-UNet, and SegFormer have demonstrated superior segmentation performance across various medical imaging tasks (Khan et al., 2025). The hybrid CNN-Transformer models leverage the advantages of both CNNs and Transformers to improve segmentation results by integrating contextual information alongside the local features extracted by CNNs. But these Transformer models are limited in their use in resource-constrained healthcare settings due to their large datasets, high computational requirements, and extensive training time.

3. FEDERATED LEARNING ARCHITECTURES

Federated Learning (FL) is a decentralized machine learning approach in which multiple healthcare institutions can train AI models collaboratively without sharing their sensitive patient data. This not only addresses privacy concerns about centralized data storage but also enables safe collaborative learning between hospitals and medical research centers. A number of federated learning architectures have been proposed, depending on the needs of data distribution and institutional collaboration. In the healthcare sector, where multiple institutions have datasets with similar feature spaces but varying patient sample sizes, the most prevalent method of horizontal Federated Learning is used. Vertical Federated Learning is for scenarios in which institutions have overlapping patient populations but different feature attributes, and Hybrid Federated Learning incorporates elements of both approaches to enable more complex collaboration scenarios.

The typical federated training process involves performing local training at each institution and then collecting the model parameters from all institutions via a central aggregator (Federated Averaging, or FedAvg) (Khan et al., 2025). Advanced architectures such as FedProx, FedAvg, and asynchronous federated optimization methods have been proposed to improve convergence stability and address the non-IID data challenge (Nemati et al., 2025). The concept of cross-silo federated learning is well-suited to healthcare, as hospitals typically have fixed infrastructure and a limited participant base. However, federated learning architectures come with drawbacks such as communication overhead, client heterogeneity, synchronization latency, scalability issues, and performance degradation when faced with highly diverse clinical datasets.

4. PRIVACY-PRESERVING AI TECHNIQUES

In the healthcare sector, where patient information is highly sensitive, maintaining privacy is crucial for AI-driven applications. In the healthcare industry, AI applications handle sensitive patient data, making privacy preservation a significant challenge. In a federated learning setup, privacy-preserving techniques are incorporated to safeguard information at the local level and to secure interactions among the healthcare institutions. One of the most popular methods is Differential Privacy, in which statistical noise is carefully tuned and added to the gradients or model parameters before sharing, thereby obscuring individual patient data (Altschuler & Talwar, 2022). This approach allows for a decrease in the likelihood of privacy loss while preserving the model's usability and segmentation accuracy.

Another significant technique that allows the aggregation of multiple client updates to the model in an encrypted manner, without disclosing individual institutional contributions, is also important. Another advantage of Homomorphic Encryption is that it enables mathematical operations to be performed directly on encrypted data, so the data does not need to be decrypted for computation (Mollakuqe et al., 2024). Collaborative computations can also be performed using Secure Multi-Party Computation (SMPC) without exposing participating institutions' sensitive information to one another. These techniques can be used to protect federated systems from model inversion attacks, membership inference attacks, and gradient leakage attacks. The adoption of robust privacy-preserving techniques can, however, introduce computational complexity, communication costs, training latency, and resource usage, making deployment in large-scale, real-world healthcare settings more challenging.

5. CHALLENGES IN MULTI-INSTITUTIONAL HEALTHCARE DATASETS

Multi-institutional healthcare datasets come with their own set of problems that affect the performance and trustworthiness of AI models for medical image segmentation. A key problem is that the data

distribution is heterogeneous, also known as non-identically distributed (non-IID) data. The imaging devices, image acquisition protocols, image quality, annotations, patient characteristics, and other parameters used in the collection of medical images are not uniform across different hospitals (Zholshybek & Bastarbekova, 2025). These differences reduce the models' ability to generalize their segmentation and can lead to inconsistent segmentation performance across institutions.

One major problem that is still to be addressed is interoperability between health care systems. Each hospital has its own data storage, clinical standards, and information systems, making it difficult to collaborate on data use. Another critical issue in a federated learning system is communication overhead, as the continuous exchange of model information between clients and the server can introduce delays and additional computational load. Federated training also becomes more complex because some hospitals may have limited computational infrastructure or insufficient local datasets. Further, protecting privacy and complying with regulations while allowing collaborative learning remains an important challenge. Even in federated communication, information may be leaked due to security issues such as gradient leakage and inference attacks. All of these difficulties underscore the necessity of powerful, scalable, and privacy-preserving systems that can effectively handle diverse healthcare settings.

6. RESEARCH GAPS

While significant progress has been made in medical image segmentation and federated learning, several key research challenges remain unaddressed. Previous studies on medical image segmentation have primarily focused on CNN-based architectures to enhance segmentation accuracy, but few studies have explored how to incorporate advanced Transformer-based models into federated healthcare settings. Existing federated learning methods also fail to efficiently manage the case of heterogeneous and non-identically distributed (non-IID) clinical data from a variety of health care centers. Consequently, segmentation performance may be inconsistent across hospitals or under different imaging conditions.

Further, prior research has not adequately examined the efficacy of privacy-preserving techniques (differential privacy, secure aggregation, and homomorphic encryption) for large-scale medical image segmentation tasks. Most of the research focuses on model accuracy rather than on the trade-offs among privacy protection, computational efficiency, and segmentation accuracy. Furthermore, in federated healthcare systems, challenges related to communication overhead, synchronization delays, and scalability issues remain underexplored (Sana et al., 2025). Moreover, very few studies conduct comprehensive cross-institutional validation on diverse real-world medical datasets, thereby limiting the robustness and clinical applicability of previously developed models. Therefore, there is a need to develop a new robust, scalable, privacy-preserving federated deep learning framework that incorporates advanced segmentation accuracy, secure collaborative learning, and excellent generalization across heterogeneous, multi-institutional clinical datasets.

III. PROPOSED METHODOLOGY

1. FEDERATED LEARNING FRAMEWORK

The framework utilizes a decentralized federated deep learning architecture for collaborative medical image segmentation while preserving institutional data privacy. Multiple healthcare clients independently train local models using their respective medical imaging datasets, without sharing raw patient data with a centralized server. Each client trains a U-Net segmentation model locally and periodically transmits model parameters to the global aggregation server. The global model is updated through the Federated Averaging (FedAvg) optimization strategy as follows (Vashistha et al., 2025).

Where w^{k}_{t} represents local client parameters, and n_k denotes client-specific training samples. Multiple communication rounds are performed to enhance global segmentation. Differential privacy is implemented during optimization through techniques such as gradient clipping and the addition of Gaussian noise, ensuring convergence across heterogeneous institutional distributions. The decentralized architecture reduces the risk of leakage of private medical information, thereby improving the effectiveness of the collaborative learning process for multi-institutional segmentation training.

2. MEDICAL IMAGE DATASET AND PREPROCESSING

The system uses the Brain Tumor Segmentation (BraTS2020) database for decentralized segmentation of medical images via a federated learning approach. The database contains multiple MRI scans in T1, T1ce, T2, and FLAIR modalities, accompanied by corresponding tumor masks for supervised learning tasks (Kaggle.com, 2020). The MRI scans were partitioned among different simulated institution clients for the federated optimization process. The preprocessing workflow aims to improve the accuracy and stability of optimization processes on a distributed platform. Firstly, data that did not contain significant tumor-related regions in MRI slices were discarded to reduce class imbalance. In addition, normalization was applied to ensure consistency in the pixel value distribution. All the image tensors were reshaped into a uniform size of 240×240 pixels. Furthermore, binary masks for tumor localization were obtained from the annotated data to extract relevant information. Preprocessed images and their respective masks were converted into tensors using PyTorch DataLoader.

3. SEGMENTATION MODEL

The framework integrates a U-Net-based convolutional neural network (CNN) architecture for decentralized medical image segmentation within a federated learning environment (Jiangtao et al., 2025). The segmentation model consists of an encoder-decoder architecture with multiple blocks, including ReLU activation layers, skip connections, and max-pooling operations (Nworu et al., 2022). Thisisn done topreserve low-level spatial information and high-level semantictumorr representations. The encoder path progressively extracts hierarchical feature representations from MRI images, and the decoder reconstructs the fine-grained tumor segmentation masks. Skip connections are incorporated to minimize the loss of spatial information and improve boundary localization during pixel-wise segmentation. The segmentation optimization process utilizes the Dice Loss to maximize the overlap between predicted segmentation masks and annotated tumor regions.

Where pi denotes the predicted segmentation pixels, gi denotes the ground-truth labels, and ϵ is a smoothing constant for numerical stability. To enhance privacy preservation, differential privacy mechanisms were integrated into the federated optimization process via Gaussian noise perturbation. Sensitive gradient updates were constrained and perturbed before global aggregation during local client training.

Where Clip(g, C) limits the gradient magnitude, and $N(0, \sigma^2C^2)$ represents Gaussian noise injection. This mechanism minimizes the leakage of sensitive information from local model updates during collaborative federated learning.

4. FEDERATED TRAINING AND OPTIMISATION STRATEGY

The framework adopts a cross-silo federated learning architecture for decentralized medical image segmentation across multiple healthcare institutions. Participating institutional clients collaboratively train a global segmentation model without directly sharing sensitive patient MRI datasets with the central server. The global segmentation model is distributed from the aggregation server to all participating institutional clients. Each client independently performs local optimization using its respective MRI data and subsequently transmits only the locally updated model parameters. The server then integrates these updates using the Federated Averaging (FedAvg) optimization strategy to produce an improved global segmentation model for subsequent communication rounds (Nazir & Kaleem, 2023). Multiple federated communication rounds are executed to progressively refine the global segmentation model across heterogeneous institutional distributions. The decentralized optimization strategy enables collaborative medical image segmentation and minimizes dependence on centralized resources.

5. EXPERIMENTAL SETUP AND EVALUATION METRICS

The experimental framework was implemented using Python and PyTorch within the Google Colaboratory cloud environment. The segmentation experiments were conducted using a U-Net-based CNN architecture trained under federated and differentially private federated learning settings. The federated

learning environment consisted of four simulated institutional clients representing decentralized healthcare organizations participating in collaborative segmentation training. The MRI images were preprocessed through normalisation, tensor conversion, and binary mask generation before training. The dataset was partitioned into decentralized subsets to simulate heterogeneous, multi-institutional distributions for federated optimization. The local client models were trained using the Adam optimizer with iterative backpropagation-based parameter updates. The segmentation performance was evaluated using the Dice Similarity Coefficient (DSC) and Intersection over Union (IoU), which measure the overlap between the predicted tumor region and the ground-truth segmentation masks (Cai et al., 2024). The Dice Score and IoU were calculated by using the formula below.

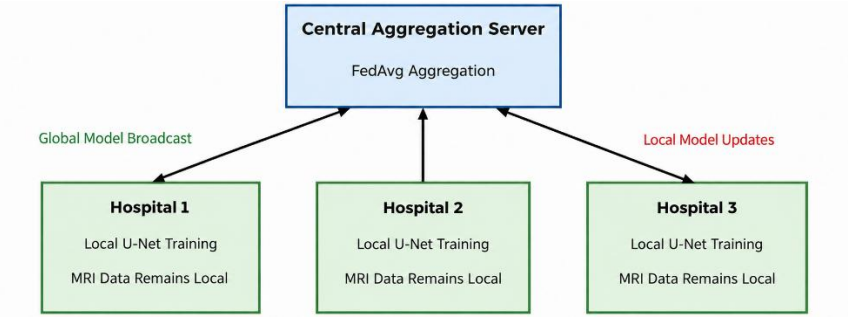
$$DSC = \frac{2TP}{2TP + FP + FN}$$

$$IoU = \frac{TP}{TP + FP + FN}$$

IV. SYSTEM DESIGN AND IMPLEMENTATION

1. SYSTEM ARCHITECTURE DESIGN

The suggested architecture for federated learning was designed as a decentralized cross-silo system for collaborative image segmentation of tumor regions across multiple healthcare institutions. It involves institutional clients, locally trained segmentations, a central aggregation server, and optimization modules for privacy preservation. The local client does not share any sensitive patient information and uses a locally stored dataset. Local convolutional neural networks based on a U-Net architecture are employed at each client for tumor region segmentation and feature extraction. Firstly, a global segmentation model is distributed by the central server to each institutional client. Local optimization of the model is performed via iterative decentralized learning with backpropagation on MRI data. Following the optimization process, the local model is sent to the central server and synchronized using the Federated Averaging (FedAvg) technique. Gradient clipping and Gaussian noise addition are used as differential privacy techniques to secure communication while optimizing federated learning models (Xiao et al., 2023).



(Source: self-developed).

FIGURE 1. Federated learning architecture for medical image segmentation.

2. CLIENT–SERVER COMMUNICATION PROCESS

The client-server interaction protocol was developed to enable decentralized, collaborative learning among the involved health care institutions and the central aggregation server. First, the global segmentation model is distributed from the central server to all institutional clients taking part in the federated learning framework. Local training on individual MRI datasets takes place at each client, and the model parameters are optimized iteratively. Then, the server collects the local updates and completes model synchronization using the FedAvg optimization algorithm. The global segmentation model is redistributed back to the institutional clients

3. FEDERATED MODEL AGGREGATION PROCESS

Federated model aggregation was performed using the Federated Averaging (FedAvg) optimization algorithm, enabling federated learning to train the segmentation task across all institutions in a decentralized manner (Khan et al., 2025). Once the local training process is complete, each institution sends optimized model parameters to the central aggregation server, which averages them according to their dataset sizes. The iterative optimization process progressively improves segmentation generalization across heterogeneous institutional MRI distributions and reduces dependence on centralized data.

4. PRIVACY AND SECURITY IMPLEMENTATION

The design framework integrates federated learning and differential privacy to protect patient privacy in collaborative medical image segmentation tasks. In this context, patients' medical imaging datasets are stored locally within participating health care organizations, without any centralized transfer of data for model training. Local clients exchange locally optimized model parameters in each round of federated communication, rather than directly sending raw patient images. Privacy protection was further improved by using differential privacy methods, such as gradient clipping and adding Gaussian noise, to optimize locally before transmitting parameters. This protects sensitive gradient information to prevent the reconstruction of patients' private data during parameter transmission.

5. MODEL IMPLEMENTATION ENVIRONMENT

The above-mentioned federated learning medical image segmentation architecture was implemented in Python on Google Colaboratory, leveraging a GPU-based cloud computing environment. The design of deep learning models and the process of federated optimization were implemented using the PyTorch library because of its fast tensor computations and its ability to handle dynamically defined neural networks. All processes associated with MRI image processing, together with tensor transformations, were carried out using the NumPy and h5py libraries, whereas visualization and evaluation were performed using the Matplotlib library (Andersen, 2022). Thus, the computing environment used enabled GPU computations that made federated healthcare experiments computationally feasible.

V. RESULTS AND ANALYSIS

1. SEGMENTATION PERFORMANCE RESULTS

The segmentation performance was evaluated using the DSC and IoU metrics to measure the overlap between predicted tumor masks and ground-truth segmentation labels. The centralised U-Net segmentation model achieved the highest segmentation performance with an average Dice Score of 0.3628 and an IoU of 0.2464. The results show stable convergence of optimization and improved tumor boundary localization. The federated learning framework achieved an average Dice Score of 0.3295 and an IoU score of 0.2176. This indicates that decentralized collaborative learning successfully preserved segmentation capability. The minimal performance gap confirmed the effectiveness of the FedAvg-based aggregation mechanism for cross-institutional segmentation learning. The DP-federated model achieved lower segmentation performance, with Dice and IoU scores of 0.04408 and 0.0211, respectively. The segmentation visualization results further confirmed that the centralized and federated models successfully captured the major tumor regions (Darzidehkalani et al., 2022). The results demonstrated that federated learning achieved competitive segmentation performance while significantly improving the preservation of decentralized medical data privacy.

Table 1. Dice and iou comparison table.

Model	Average Dice Score	Average IoU Score
Centralized U-Net	0.3628	0.2464
Federated U-Net	0.3295	0.2176
DP-Federated U-Net	0.0408	0.0211

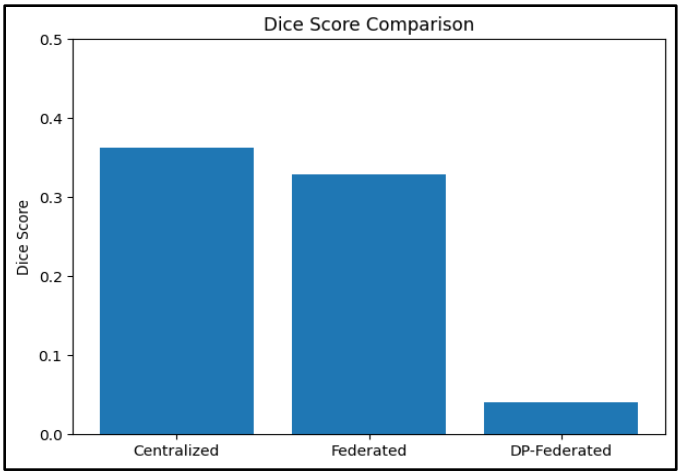


FIGURE 2. Dice score comparison.

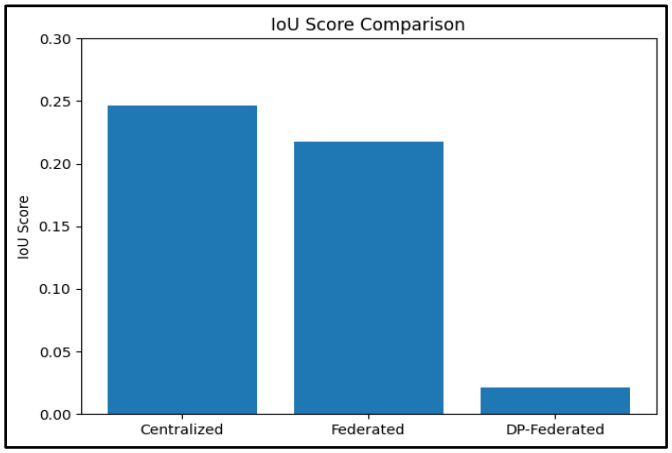
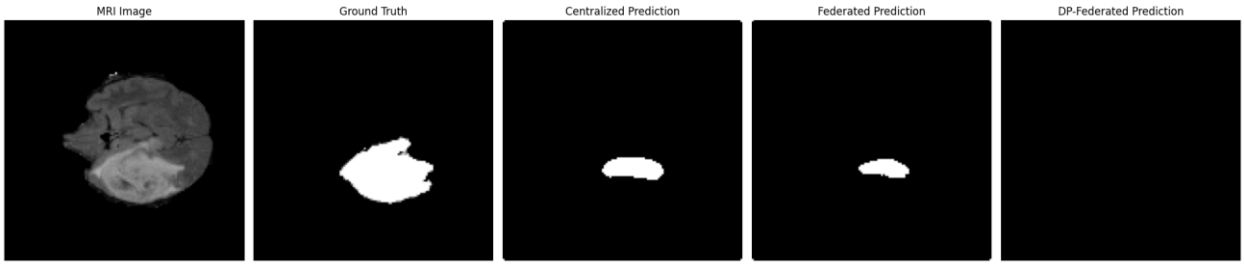


FIGURE 3. IoU Score comparison.

2. COMPARATIVE ANALYSIS WITH CENTRALISED MODELS

A comparison of the three learning paradigms used for segmentation is presented in Table 5.2. In terms of medical image segmentation, the centralized U-Net model achieved the highest Dice and IoU scores, with mean values of 0.3628 and 0.2464, respectively. The federated learning paradigm recorded average Dice Score and IoU Score of 0.3295 and 0.2176, respectively, showing a minor decrease in performance compared with the centralized learning paradigm. This small difference in performance indicated the viability of decentralized collaborative optimization using the FedAvg algorithm. Even though the MRI dataset was not shared centrally, the federated learning paradigm demonstrated satisfactory generalization to the distributed institutional datasets. The DP-Federated paradigm exhibited loDicedice and IoU scores of 0.0408 and 0.,0211 respectively. The segmentation performance comparison between centralised, federated, and DP federated learning architectures is presented in Figure 5.3. Segmentation accuracy was higher for the centralized U-Net architecture, as it achieved the best overlap with the ground-truth mask. Furthermore, the federated learning architecture could detect the main tumor area, but the overlap with the ground-truth mask was somewhat smaller than that obtained using the other two approaches (Xin et al., 2024). Conversely, the use of privacy-preserving techniques in the federated DP learning architecture led to empty masks due to the added noise.



FIGURECentralizedised vs federatedDPs dp segmentation outputs.

Table 2. Comparative performance analysis.

Model	Segmentation Accuracy	Privacy Preservation	Optimization Stability	Data Sharing Requirement
Centralized U-Net	High	Low	Stable	Full MRI Sharing
Federated U-Net	Moderate–High	Moderate	Stable	Parameter Sharing Only
DP-Federated U-Net	Low	Strong	Unstable	No Direct Data Sharing

3. CROSS-INSTITUTIONAL GENERALISATION ANALYSIS

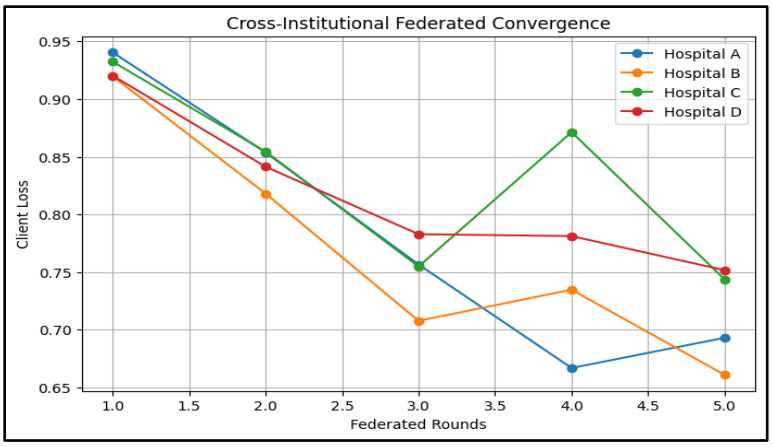


FIGURE 5. Cross-institutional federated convergence graph.

Table 3. Client-wise federated loss values across rounds.

Federated Round	Hospital A	Hospital B	Hospital C	Hospital D
Round 1	0.9405	0.9198	0.9324	0.9203
Round 2	0.8537	0.8180	0.8544	0.8413
Round 3	0.7565	0.7077	0.7548	0.7828
Round 4	0.6668	0.7347	0.8713	0.7812
Round 5	0.6930	0.6609	0.7437	0.7516

The cross-institutional generalization analysis assessed the convergence performance of the federated learning framework across 4 healthcare organizations. In this context, Table 5.3 and the figure show that all hospitals achieved a gradual reduction in loss over 5 federated communication rounds, indicating stable collaboration and optimization. Primarily, all clients recorded high loss values in Round 1, ranging from 0.9198 to 0.9405, as the global segmentation model had not yet optimized effectively. Therefore, constant parameter aggregation via the FedAvg algorithm enhanced model learning in subsequent rounds. Hospital

A's loss value decreased from 0.9405 to 0.6930, and Hospital B had the lowest final loss value of 0.6609. Likewise, decreasing patterns were observed at Hospital C and Hospital D, with final losses of 0.7437 and 0.7516, respectively. These results confirmed that the federated framework can effectively transfer tumor segmentation knowledge from multiple datasets across different institutions while avoiding direct sharing of MRI data (Albalawi et al., 2024). As a result, the proposed federated architecture exhibited good collaborative learning capability and adaptability for multi-institution medical image segmentation tasks.

4. PRIVACY PRESERVATION AND SECURITY ANALYSIS

Table 4. Privacy vs performance comparison.

Model	Dice Score	IoU Score	Privacy Level	Security Strength
Centralized U-Net	0.3628	0.2464	Low	Vulnerable to centralized leakage
Federated U-Net	0.3295	0.2176	Moderate	Reduced data exposure
DP-Federated U-Net	0.0408	0.0211	Strong	High resistance to inference attacks

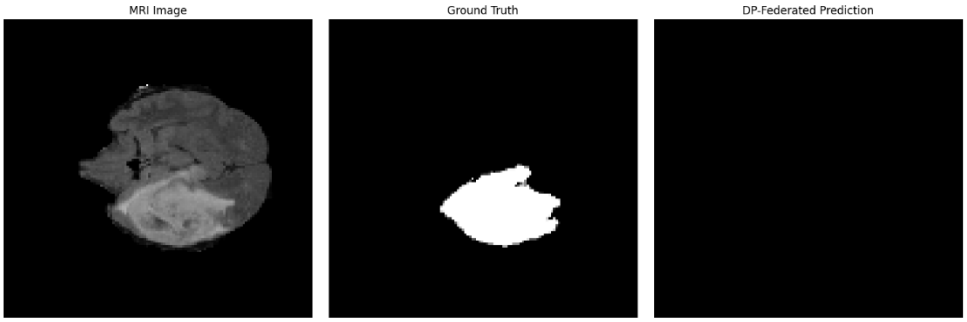


FIGURE 6. DP-Federated prediction visualisation.

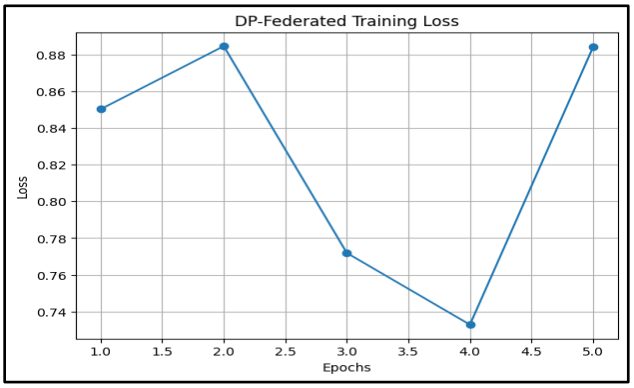


FIGURE7. DP Training loss curve.

The privacy and security analysis compared the trade-off between segmentation accuracy and patient data protection across centralized, federated, and DP-federated learning methods. In this context, Table 5.4 shows that the centralized U-Net achieved the maximum Dice Score (0.3628) and IoU Score (0.2464), as all MRI data were available centrally during optimization. Therefore, this approach offered less privacy protection and greater vulnerability to centralized data leakage. In addition, the federated U-Net architecture achieved slightly lower Dice and IoU scores of nearly 0.3295 and 0.2176, respectively, while significantly enhancing data security. Patient MRI datasets remained stored locally within the healthcare organizations, and only model parameters were exchanged during training (Aggarwal et al., 2023). This primarily minimized sensitive data exposure and improved collaborative privacy preservation. The DP-Federated U-Net technique achieved the highest level of privacy protection by implementing gradient clipping on the model and adding Gaussian noise to the training data. The results in Figures 5.6 and 5.7 showed that privacy-

preserving optimization improved resistance to inference attacks and prevented the leakage of sensitive information. But when too much noise was injected into the data, it made optimization unstable and reduced segmentation accuracy, leading to Dice and IoU scores of 0.0408 and 0.0211, respectively.

5. ROBUSTNESS AND SCALABILITY EVALUATION

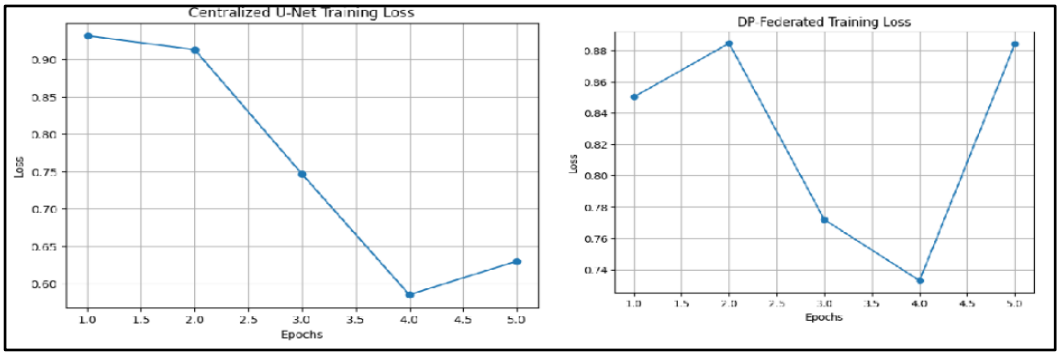


FIGURE 8. Centralized and DP training loss curves.

Table 5. Scalability and robustness comparison.

Evaluation Factor	Centralized U-Net	Federated U-Net	DP-Federated U-Net
Training Stability	High	Moderate–High	Moderate
Cross-Institution Generalisation	Limited	Strong	Moderate
Computational Complexity	Moderate	High	Very High
Scalability Across Institutions	Limited	High	High
Communication Overhead	Low	Moderate	High
Privacy Preservation	Low	Moderate	Strong
Segmentation Robustness	High	Moderate–High	Low

The scalability evaluation mostly assessed the overall operational efficiency of the segmentation architectures in decentralized healthcare environments. In this context, Table 5.5 and Figure 5.8 show that the centralized U-Net model has achieved high segmentation robustness and stable training convergence because of complete access to centralized MRI datasets. In addition, the DP-Federated U-Net has achieved maximum privacy protection and scalability, thus requiring additional computational resources for privacy-preserving optimization. Differential privacy mechanisms incur communication overhead and reduce convergence stability, leading to less segmentation robustness (Yang et al., 2025).

VI. DISCUSSION

1. INTERPRETATION OF FINDINGS

The results of the study showed that federated deep learning could offer a viable trade-off between medical image segmentation and privacy protection among distributed health data. The U-Net model with the most centralized structure (the highest Dice Score (0.3628) and IoU Score (0.2464)) demonstrated better localization of tumor boundaries and greater overlap in segmentation, with direct access to the full MRI dataset. Nevertheless, the federated U-Net model achieved comparable results, with Dice and IoU scores of 0.3295 and 0.2176, respectively. It demonstrates that decentralized collaborative learning can maintain the ability to perform segmentation without necessarily sharing sensitive patient data. These results suggest that the FedAvg optimization system was able to transfer knowledge of tumor segmentation across institutional datasets and that the optimization was stable.

They are consistent with the article by Ullah et al. (2023), which found that federated U-Net models achieve the same segmentation fidelity as a centralized model and can improve privacy for medical data. In their article, they emphasized that this parameter-sharing process allows the models to learn robust tumor features while benefiting from distributed clinical information without exposing raw patient image data.

Similarly, these results are in agreement with the publication by Chowdhury & Kudapa (2024), who also confirmed that federated learning reduces the risk of centralized data leaks and enables institutions to comply with healthcare privacy regulations. The current investigation also established that parameter sharing reduced sensitive MRI exposure without compromising the convergence across hospitals. In the current study, the DP-Federated model yielded much lower Dice and IoU scores due to Gaussian noise and gradient clipping that decreased the capacity of the model in delineating fine tumor edges.

2. CLINICAL SIGNIFICANCE

The clinical implications of the study are that this research can be used to achieve safe and effective medical image segmentation across multiple healthcare facilities without necessarily sharing patient information. The federated U-Net model has shown how hospitals can jointly train deep learning models to segment tumours, maintaining patient confidentiality and in accordance with healthcare privacy laws. This is crucial in clinical settings where sensitive MRI datasets cannot be transferred centrally due to ethical, legal, and institutional limitations. The suggested federated learning framework has the potential to enhance clinical decision-making by enabling healthcare organizations to access a variety of multi-institutional datasets, thereby improving the overall generalization of tumor segmentation systems. The consistent convergence across all involved hospitals suggests that collaborative learning could facilitate scalable implementation in distributed healthcare settings. Moreover, the use of differential privacy enhanced the defense against inference attacks and unauthorized access to patient data, which is essential for trustworthy AI use in healthcare systems.

Even though the DP-Federated model reduced segmentation performance due to noise injection, it nonetheless demonstrated the effectiveness of privacy-sensitive medical AI systems in practice. Consequently, the paper offers useful findings that federated deep learning can help enable future intelligent diagnostic and telemedicine systems, as well as collaborative health analytics, without compromising data protection, institutional autonomy, or patient confidentiality.

3. ADVANTAGES AND LIMITATIONS

The idea of privacy-preserving federated deep learning is beneficial for medical image segmentation in several ways. It also helps organizations train models together while maintaining the privacy of their MRI data, potentially helping them adhere to privacy regulations and laws. The federated U-Net matched the segmentation performance of the centralized model, with stable optimization and good cross-institutional generalization. In addition, the framework showed its validity when applied to a distributed health care environment. However, some limitations were identified. The Federated model had lower segmentation performance than the centralized route. Dice and IoU scores dropped significantly in the differential privacy setting due to the noisy samples introduced during training. Also, federated learning requires more communication overhead, additional computation, and longer training times, potentially limiting its deployability in healthcare environments with limited resources.

4. ETHICAL AND PRIVACY CONSIDERATIONS

Due to the presence of patient MRI scans containing sensitive personal information, ethical and privacy concerns are paramount in medical applications of AI. However, these concerns were addressed with a proposed federated learning (FL) framework, which allowed patient data to remain within local healthcare institutions, with only model parameters sent during training. This helped minimize the chances of unauthorized access, data breaches, and privacy violations. Furthermore, the integration of differential privacy enhanced protection against inference attacks by introducing noise into model updates, making it more difficult to reconstruct patient information. The framework is also built to comply with guidelines and requirements of healthcare data protection and medical research ethics. However, it is crucial to achieve this equilibrium between privacy protection and model accuracy, as robust privacy measures can impact the clinical performance and diagnostic reliability of models.

VII. CONCLUSION AND FUTURE WORK

1. CONCLUSION

In this study, a Privacy-preserving Federated deep learning framework for robust medical image segmentation over multi-institutional clinical datasets has been presented. The key goal was to assess the ability of federated learning to maintain performance in tumor segmentation without centralizing data and without revealing patients' sensitive information. The three learning paradigms assessed were a centralized U-Net, a federated U-Net,t and a DP-Federated U-Net. Experimental results showed that the Centralized U-Net achieved the best segmentation, with an average Dice Score of 0.3628 and an average IoU Score of 0.2464. The federated U-Net achieved competitive performance with Dice 0.3295 and IoU 0.2176. Results from these studies showed that collaborative learning with federated optimization can be used to learn tumor characteristics from distributed MRI datasets without relinquishing data ownership to other institutions. Results indicated a relatively small difference between centralized and federated systems, thereby demonstrating the feasibility of decentralized medical image analysis systems. Also, stable convergence across all participating hospitals was shown in the cross-institutional evaluation. Loss consistently decreased during federated communication rounds, demonstrating the positive impact of the FedAvg aggregation strategy on collaborative model optimization. Federated learning can be used to enable large-scale collaborations among healthcare systems that have limited ability to share data directly due to privacy, legal, or organizational reasons. The work also assessed the effects of differential privacy on the segmentation performance and security. The DP-Federated U-Net offered the highest level of privacy (while preserving its segmentation accuracy), but too much noise injection caused major drops in segmentation accuracy. This result underscored the crucial balance needed between privacy protection and predictive accuracy in healthcare AI systems. The results showed that federated deep learning is a potential approach for image segmentation in medical applications while maintaining privacy. The proposed framework achieved a good balance among segmentation effectiveness, scalability, and data protection across multiple institutions.

2. FUTURE WORK

Future studies may explore enhanced segmentation accuracy for privacy-protection mechanisms in federated learning architectures without compromising security. The DP-Federated U-Net model obtainedlowers Dice and IoU scores due toexcessiveh Gaussian noise and gradient perturbatio, which disturbs theoptimizationn stability and thetumorr boundary localization. Thus, future research may explore adaptive DP methods that adjust the trade-off between privacy and segmentation accuracy as optimization progresses. The performance may be further enhanced by using advanced aggregation algorithms beyond FedAvg, such as attention-based or personalized federated optimization algorithms, to achieve better convergence on heterogeneous institutional data. Further improvements could be made by using larger multimodal medical imaging datasets from real healthcare institutions rather than simulated client distributions.

Future frameworks could incorporate transformer-based segmentation or combined CNN-transformer architectures to improve feature extraction and tumor localization accuracy. Also, communication-efficient federated learning techniques can be investigated to mitigate the high computational complexity, communication overhead, and training latency in distributed healthcare environments. Security can also be strengthened by integrating secure multi-party computation and homomorphic encryption mechanisms to better protect against inference attacks and adversarial threats. Finally, the proposed framework could be applied in real-world clinical settings to assess its scalability, reliability, and usability for intelligent diagnostic and telemedicine systems in the future.

Author Contributions

The author conducted the conceptualization, methodology, data analysis, investigation, writing, review, editing, and final approval of the manuscript.

Funding

This research received no external funding.

Data Availability

The dataset will be available from the author upon reasonable request.

Conflicts of Interest

The author declares no conflict of interest.

REFERENCES

1. Abueed, O., Wang, Y., & Khasawneh, M. (2025). A Systematic Review of U-Net Optimizations: Advancing Tumour Segmentation in Medical Imaging. *IET Image Processing*, *19*(1), e70203. https://doi.org/10.1049/ipr2.70203
2. Aggarwal, K., Manso Jimeno, M., Ravi, K. S., Gonzalez, G., & Geethanath, S. (2023). Developing and deploying deep learning models in brain magnetic resonance imaging: a review. *NMR in Biomedicine*, *36*(12), e5014. https://arxiv.org/pdf/2301.01241
3. Albalawi, E., TR, M., Thakur, A., Kumar, V. V., Gupta, M., Khan, S. B., & Almusharraf, A. (2024). Integrated approach of federated learning with transfer learning for classification and diagnosis of brain tumor. *BMC Medical Imaging*, *24*(1), 110. https://doi.org/10.1186/s12880-024-01261-0
4. Altschuler, J., & Talwar, K. (2022). Privacy of noisy stochastic gradient descent: More iterations without more privacy loss. *Advances in Neural Information Processing Systems*, *35*, 3788- 3800. https://proceedings.neurips.cc/paper_files/paper/2022/file/18561617ca0b4ffa293166b3186e04b0-Paper-Conference.pdf
5. Andersen, E. (2022). Imagedata: A Python library to handle medical image data in NumPy array subclass Series. Journal of Open Source Software, 7(73), 4133. https://doi.org/10.21105/joss.04133
6. Azad, R., Aghdam, E. K., Rauland, A., Jia, Y., Avval, A. H., Bozorgpour, A., ... & Merhof, D. (2024). Medical image segmentation review: The success of U-Net. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *46*(12), 10076-10095. https://doi.org/10.1109/TPAMI.2024.3435571
7. Barbaria, S., Jemai, A., Ceylan, H. İ., Muntean, R. I., Dergaa, I., & Boussi Rahmouni, H. (2025, October). Advancing compliance with HIPAA and GDPR in healthcare: a blockchain-based strategy for secure data exchange in clinical research involving private health information. In *Healthcare* (Vol. 13, No. 20, p. 2594). MDPI. https://doi.org/10.3390/healthcare13202594
8. Cai, L., Li, Y., Fan, X., Song, K., Wang, R., & Lei, W. (2024). Low-contrast-enhanced contrastive learning for semi-supervised endoscopic image segmentation. arXiv preprint arXiv:2412.02314. https://arxiv.org/pdf/2412.02314
9. Chowdhury, T. K., & Kudapa, S. P. (2024). Federated Learning Models for Privacy-Preserving Data Sharing And Secure Analytics In Healthcare Industry. International Journal of Business and Economics Insights, 4(4), 91-133.https://doi.org/10.63125/c2dzn006
10. Darzidehkalani, E., Ghasemi-Rad, M., & Van Ooijen, P. M. A. (2022). Federated learning in medical imaging: part I: toward multicentral health care ecosystems. Journal of the american college of radiology, 19(8), 969-974. https://doi.org/10.1016/j.jacr.2022.03.015
11. Fareed, S., Yi, D., Hussain, B., Uddin, S., Arif, A., & Tajoor, A. N. (2025). Fedsegnet: A federated learning framework for 3d medical image segmentation. *International Journal of Ethical AI Application*, *1*(2), 30-46. https://doi.org/10.64229/ttnjjp90
12. Gu, X., Sabrina, F., Fan, Z., & Sohail, S. (2023). A review of privacy enhancement methods for federated learning in healthcare systems. *International Journal of Environmental Research and Public Health*, *20*(15), 6539. https://doi.org/10.3390/ijerph20156539
13. Hussain, D., Al-Masni, M. A., Aslam, M., Sadeghi-Niaraki, A., Hussain, J., Gu, Y. H., & Naqvi, R. A. (2024). Revolutionizing tumor detection and classification in multimodality imaging based on deep learning approaches: Methods, applications, and limitations. *Journal of X-Ray Science and Technology*, *32*(4), 857-911. https://doi.org/10.3233/XST-230429
14. Hussain, S., Mubeen, I., Ullah, N., Shah, S. S. U. D., Khan, B. A., Zahoor, M., ... & Sultan, M. A. (2022). Modern diagnostic imaging technique applications and risk factors in the medical field: a review. *BioMed research international*, *2022*(1), 5164970. https://doi.org/10.1155/2022/5164970
15. Jiangtao, W., Ruhaiyem, N. I. R., & Panpan, F. (2025). A comprehensive review of U-Net and its variants: advances and applications in medical image segmentation. IET Image Processing, 19(1), e70019. https://doi.org/10.1049/ipr2.70019
16. Kaggle.com (2020). Brain Tumor Segmentation(BraTS2020). (2020). Retrieved from www.kaggle.com website: https://www.kaggle.com/datasets/awsaf49/brats2020-training-data
17. Khan, A., Rauf, Z., Khan, A. R., Rathore, S., Khan, S. H., Shah, N., ... & Gwak, J. (2025). A recent survey of vision transformers for medical image segmentation. *IEEE Access*. https://doi.org/10.1109/ACCESS.2025.3618215
18. Khan, N., Nisar, S., Khan, M. A., Rehman, Y. A. U., Noor, F., & Barb, G. (2025). Optimizing federated learning with aggregation strategies: A comprehensive survey. *IEEE Open Journal of the Computer Society*. https://doi.org/10.1109/OJCS.2025.3590102

19. Mollakuqe, E., Parduzi, A., Rexhepi, S., Dimitrova, V., Jakupi, S., Muharremi, R., ... & Qarkaxhija, J. (2024). Applications of homomorphic encryption in secure computation. *Open Research Europe*, *4*, 10-12688. https://dx.doi.org/10.12688/openreseurope.18052.1
20. Myakala, P. K., Jonnalagadda, A. K., & Bura, C. (2024). Federated learning and data privacy: A review of challenges and opportunities. *International Journal of Research Publication and Reviews*, *5*(12), 10-55248. https://doi.org/10.55248/gengpi.5.1224.3512
21. Nazir, S., & Kaleem, M. (2023). Federated learning for medical image analysis with deep neural networks. Diagnostics, 13(9), 1532. https://doi.org/10.3390/diagnostics13091532
22. Nemati, A. A., Sadreddini, M. H., & Mahdizade, M. (2025). Comparative Federated Algorithms for Solving Non-IID Data Challenges. *Transactions on Soft Computing*, *1*(1), 27-35. https://doi.org/10.48314/tsc.v1i1.35
23. Nworu, C. C., Ekpenyong, J. E., Chisimkwuo, J., Okwara, G., Agwu, O. J., & Onyeukwu, N. C. (2022). the effects of modified ReLU activation functions in image classification. J Biomed. Eng. Med. Dev., 7, 237. DOI: 10.35248/ 2475 7586.22.07.237
24. Salman, H. A., & Kalakech, A. (2024). Image enhancement using convolutional neural networks. *Babylonian Journal of Machine Learning*, *2024*, 30-47. https://doi.org/10.58496/BJML/2024/003
25. Sana, T. Z., Abdulla, S., Das, A., Nag, A., Hassan, M. M., Fiza, Z. Z., ... & Kabir, S. R. R. (2025). Advancing federated learning: A systematic literature review of methods, challenges, and applications. *IEEE Access*. https://doi.org/10.1109/ACCESS.2025.3605165
26. Sistaninejhad, B., Rasi, H., & Nayeri, P. (2023). A review paper about deep learning for medical image analysis. *Computational and Mathematical Methods in Medicine*, *2023*(1), 7091301. https://doi.org/10.1155/2023/7091301
27. Ullah, F., Nadeem, M., Abrar, M., Amin, F., Salam, A., & Khan, S. (2023). Enhancing brain tumor segmentation accuracy through scalable federated learning with advanced data privacy and security measures. Mathematics, 11(19), 4189.https://doi.org/10.3390/math11194189
28. Vashistha, P., Vashistha, D., Saxena, V. P., Kumhar, M., Bhatia, J., Gupta, R., ... & Tanwar, S. (2025). Federated Averaging Optimization for Efficient Skin Cancer Image Analysis. *Procedia Computer Science*, *258*, 3794-3803. DOI: 10.1016/j.procs.2025.04.634
29. Veiga-Canuto, D., L. Cerdá Alberich, C. Sangüesa Nebot, Martínez, B., Pötschger, U., Michela Gabelloni, …Martí-Bonmatí, L. (2022). Comparative Multicentric Evaluation of Inter-Observer Variability in Manual and Automatic Segmentation of Neuroblastic Tumors in Magnetic Resonance Images. *Cancers*, *14*(15), 3648–3648. https://doi.org/10.3390/cancers14153648
30. Xiao, H., Wan, J., & Devadas, S. (2023, November). Geometry of sensitivity: Twice sampling and hybrid clipping in differential privacy with optimal Gaussian noise and application to deep learning. In Proceedings of the 2023 ACM SIGSAC Conference on Computer and Communications Security (pp. 2636-2650). https://doi.org/10.1145/3576915.3623142
31. Xin, W., Jiaqian, L., Xueshuang, D., Haoji, Z., & Lianshan, S. (2024). A survey of differential privacy techniques for federated learning. IEEE Access, 13, 6539-6555. Digital Object Identifier 10.1109/ACCESS.2024.3523909
32. Yang, Z., Yan, X., Chen, G., & Tian, X. (2025). Adaptive Differential Privacy for Satellite Image Recognition with Convergence-Guaranteed Optimization. *Electronics*, *14*(18), 3680. https://doi.org/10.3390/electronics14183680
33. Zhang, K., Wang, J., Wang, W., Zeng, T., Li, P., Wang, X., & Zhang, T. (2025). Federated learning with heterogeneous data and models based on global decision boundary distillation. *Journal of King Saud University - Computer and Information Sciences*, *37*(5). https://doi.org/10.1007/s44443-025-00097-0 https://doi.org/10.1007/s44443-025-00097-0
34. Zholshybek, N., & Bastarbekova, L. (2025). Harmonizing multicenter quantitative imaging data: sources of variability, statistical solutions, and practical workflows in CT and MRI. *Exploration of Digital Health Technologies*, *4*.