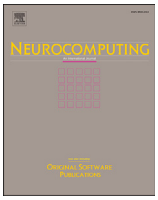
Neurocomputing 677 (2026) 133142

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom



Towards robust neurocomputing model in efficient federated brain tumour segmentation with sparsification and weights clustering

Asaf Raza [a,1], Ciro Benito Raggio [b,1], Antonella Guzzo [a,*], Maria Francesca Spadea [b], Giancarlo Fortino [a]

[a] DIMES–Department of Informatics, Modeling, Electronics and Systems, University of Calabria, Via P. Bucci, Rende, 87036, Cosenza, Italy
[b] Institute of Biomedical Engineering, Karlsruhe Institute of Technology, Fritz-Haber-Weg 1, Karlsruhe, 76131, Baden-Württemberg, Germany

HIGHLIGHTS

- Analyse the impact of FedWSOComp, an integrated strategy combining top-k sparsification, quantization, and entropy-based encoding.
- The performance of 18 different configurations (including IID and non-IID) was evaluated systematically.
- High retention (60%) with fine quantization (64 clusters) optimizes performance.

ARTICLE INFO

Communicated by N. Zeng

Keywords:
Brain tumour segmentation
Communication-efficient federated learning
Medical imaging
Quantization

ABSTRACT

Brain tumour segmentation is a key application of AI in neuroimaging. Recently, federated learning (FL) has emerged as a strategic and increasingly relevant paradigm in neural computing due to its ability to address key challenges in large-scale neural network training, such as data access, privacy, collaborative learning, and model robustness. However, its adoption is currently hindered by high communication costs and the heterogeneity of client data. In this study, we investigated an efficient FL framework for brain tumour segmentation based on communication-aware optimization. We evaluated FedWSOComp, which integrates sparsification, quantization, and entropy-based encoding, in combination with a 3D U-Net architecture under both homogeneous and heterogeneous data distributions. The multi-institutional FeTS 2024 dataset was employed and partitioned into independent and identically distributed (IID) and non-IID settings, with an independent test set of 67 patients. An overall of 18 configurations combined sparsification rates and quantization levels. Performance was measured using Dice Similarity Coefficient (DSC) and 95th percentile Hausdorff Distance (HD95). Experimental results demonstrated that aggressive compression caused severe degradation in segmentation quality, with HD95 exceeding 60 mm. In contrast, higher retention with finer quantization achieved the best balance between efficiency and accuracy, reaching a DSC of 0.98 ± 0.09 and HD95 of 10.40 ± 15.54 mm on the test set under non-IID conditions. The findings demonstrated that, when configured with moderate-to-fine quantization and high sparsification retention, FedWSOComp enabled accurate and communication-efficient federated brain tumour segmentation. This study provides quantitative evidence and practical guidance for the deployment of FL-based segmentation models in privacy-sensitive and bandwidth-constrained clinical settings.

1. Introduction

Brain tumour segmentation plays a vital role in identifying and delineating abnormal regions within brain imaging data, enabling accurate diagnosis and treatment planning and it is a major use case in AI for neuroimaging. Among the primary imaging modalities employed, computed tomography (CT) and magnetic resonance imaging (MRI) are widely used to detect and segment tumours [1]. However, MRI is generally preferred due to its non-invasive nature [2], and it remains one of the most fundamental and essential tools in the diagnosis and segmentation of brain tumours [3]. In recent years, the tumour segmentation field

* Corresponding author.
Email address: antonella.guzzo@unical.it (A. Guzzo).

[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.neucom.2026.133142
Received 15 December 2025; Received in revised form 9 February 2026; Accepted 20 February 2026
Available online 24 February 2026

has made substantial progress through the application of deep learning (DL), which offers the potential to automate tumour segmentation with high accuracy [4]. The U-Net architecture, originally proposed by Ronneberger et al. [5], has become a standard in DL-based segmentation tasks. This architecture and its numerous variants continue to serve as the foundation for state-of-the-art models in benchmark challenges, such as the Brain Tumour Segmentation (BraTS) challenge, where it has consistently been employed by top-performing teams. Despite their success and the higher segmentation accuracy, DL models typically require large volumes of annotated data [6], and accessing extensive datasets distributed across multiple institutions remains a significant challenge due to stringent data protection regulations, such as the GDPR and HIPAA [7]. Federated Learning (FL) has been identified as a promising approach for neural computing because it enables the training of complex neural models on distributed and sensitive data without centralizing them. In a typical FL setup, participating clients train a local model on their private data and periodically share model weight updates with a central server. The server aggregates these updates and sends back the updated global model to all clients, enabling iterative improvement [8]. This decentralised approach mitigates data privacy concerns and fosters cross-institutional collaboration in medical imaging research [9–11]. Despite its well-recognized advantages, the application of FL to 3D medical images, such as MRI, poses significant challenges due to the high communication overhead, particularly when dealing with large DL models and resource-constrained clients [10,12]. Indeed, traditional FL aggregation strategies, such as federated averaging (FedAvg) [8], typically transmit full precision model updates during each communication round, resulting in a substantial bandwidth consumption [13]. To mitigate this, lossy compression techniques—including quantization and gradient sparsification—have been explored [14].

However, these methods often suffer from reduced accuracy and limited adaptability to the heterogeneity inherent in 3D volumetric imaging data in practical settings [15]. For instance, top-k gradient sparsification effectively reduces communication costs, it introduces approximation errors that impair convergence, especially at high sparsification levels [16]. Similarly, gradient quantization methods, such as QSGD [17] and TernGrad [18], have the potential to reduce communication overhead but often distort the gradient distributions and fail to optimise the bit efficiency in the quantization, leading to degraded model performance [19]. To address this limitation, Onaizah et al. introduced FedWSOComp [20], an advanced aggregation strategy that combined deterministic k-means clustering for sparse quantization, weight magnitude-based sparsification, and Huffman encoding. This approach significantly reduced communication costs while preserving model performance. In contrast to conventional compression methods, FedWSOComp placed emphasis on near-lossless weight recovery, an essential requirement for FL in clinical applications. However, FedWSOComp has not yet been analysed in terms of the individual effects of its core components, sparsification and clustering, particularly under conditions of data heterogeneity. Analysing the impact of sparsification and clustering in FedWSOComp is essential because these mechanisms directly affect communication efficiency, optimization dynamics, and model robustness. Sparsification reduces communication overhead but may introduce bias and noise in the weight updates, influencing convergence and the quality of learned neural representations. Clustering groups clients with similar data distributions or learning behaviors, mitigating non-independent and identically distributed (IID) effects and improving aggregation stability. Based on the above considerations, studying their combined impact is therefore crucial to understanding the trade-offs between efficiency, accuracy, and scalability in federated neurocomputing systems. In this study, we address this gap by investigating the contributions of these components within FedWSOComp in the context of 3D brain tumour segmentation. We conduct our experiments using the Federated Tumour Segmentation (FeTS) 2024 challenge dataset under both IID and non-IID settings. It is important to highlight that with respect to other FL domains, in 3D brain segmentation, it is crucial to understand the individual effects of the FedWSOComp components, since brain MRI data is highly heterogeneous across institutions due to differences in scanners, protocols and patient populations. Furthermore, segmentation tasks are extremely sensitive to small errors that can alter anatomical boundaries or clinical interpretations. It is essential to understand whether communication savings from sparsification or robustness from clustering primarily drive performance, in order to ensure that models remain accurate, fair and reliable. This analysis not only strengthens scientific understanding, but is also a prerequisite for the trustworthy deployment of FL in clinical neuroimaging, where even minor performance trade-offs can have significant medical consequences. In summary, the key contributions of this study are as follows:

- We analysed the impact of quantization parameters in the FedWSOComp framework, including sparsification rate and clustering granularity, on model convergence and segmentation performance.
- By quantifying how varying compression levels affect convergence and segmentation quality in realistic heterogeneity scenarios, the study provides insights that extend beyond a single application and are broadly valuable to researchers optimizing federated neurocomputing systems.
- We validated the real-world applicability of FedWSOComp in FL scenarios by assessing the model's generalization ability on an independent test set comprising 67 patients from 3 separate institutions that were not involved in the federation.
- We evaluated the FedWSOComp framework on the recent FeTS 2024 dataset, focusing on 3D brain tumour segmentation under both IID and non-IID settings.

2. Related work

2.1. DL architectures and FL for medical image segmentation

In recent years, DL has become increasingly prominent in 3D medical image segmentation due to its ability to capture spatial continuity and mitigate inter-slice artefacts, which are common in 2D approaches. Nonetheless, building three-dimensional models typically requires substantial computational resources. This often led to the adoption of sub-volume (or patch-based) processing, which may compromise structural consistency and the representation of global features [21,22]. These limitations became even more critical in FL, where hardware constraints and communication costs further restricted the deployment of large models on client devices. To overcome such challenges, a wide range of architectural solutions has been proposed. DenseNet-based segmentation pipelines combined with ensemble learning, for instance, enhanced feature expressiveness and robustness across multimodal MRI data [23]. Similarly, lightweight convolutional modules and Dense U-Net variants offered a favourable balance between segmentation accuracy and inference speed [22]. Attention-integrated frameworks, such as SLCA-UNet, improved border delineation by focusing on salient contextual information [21]. Other hybrid encoder–decoder models sought to improve efficiency while retaining accuracy. Obayya et al. [24], for instance, proposed a nested U-Net with multi-scale skips and adapted loss functions, specifically designed to address class imbalance in low-grade glioma segmentation. Deep supervision and dual-path encoding further strengthened structural preservation at tumour boundaries [22], while attention-gated units facilitated feature fusion across tumour subregions [21].

Related advances in medical segmentation network design further highlighted the importance of global contextual modelling and receptive field enlargement. ELTS-Net extended the 3D U-Net paradigm for liver and tumour segmentation by incorporating dilated convolutions to expand the effective receptive field, together with dual attention mechanisms operating across spatial and channel dimensions. In addition, deep supervision was employed at multiple decoder levels to

constrain intermediate representations towards target regions and improve convergence [25]. Alternative paradigms for image segmentation have also been explored beyond fully supervised convolutional models. In this context, deep reinforcement learning has been applied to medical-related image segmentation by formulating region delineation as a sequential decision-making process guided by task-specific reward functions, demonstrating feasibility on gold immunochromatographic strip images segmentation [26].

All the above methods introduce important advances in improving accuracy of learning architectures but are orthogonal with respect to the communication cost. Parallel efforts focused on adapting 3D U-Nets to federated settings. Distributed training demonstrated the feasibility of privacy-preserving segmentation in FL [27], although the volumetric nature of updates significantly increased communication overhead. To alleviate this, decentralised aggregation protocols such as peer-to-peer FL improved scalability and resilience to system faults, while compression-aware strategies were introduced to manage high-dimensional MRI data [28].

2.2. Communication-efficient FL

Heterogeneity across clients further motivated the design of adaptive solutions. Zhang et al. [6] introduced client-personalised layers, enabling global consistency alongside local adaptability. Dai et al. [29] proposed FedATA, an attention-based transformer framework that weighted client contributions by relevance, thereby enhancing robustness to non-IID data. Privacy-preserving mechanisms also gained attention, with studies exploring encryption-based methods [30] and blockchain-inspired frameworks [31] for secure medical collaborations. Lightweight and efficient convolutional networks were further applied to brain cancer analysis in FL [32]. Complementary approaches included contrastive learning [33], standardised evaluation protocols [34], and the integration of transfer learning into federated pipelines, which proved particularly beneficial in cross-institutional scenarios [35,36]. Despite these advances, communication efficiency remained a central challenge. Compression methods such as sparsification and quantization were explored to reduce transmission costs. Top-k sparsification reduced the exchange of less relevant gradient updates [37,38], while weight quantization decreased bandwidth requirements by mapping parameters to lower-precision formats [39]. Adaptive schemes, including layer-wise quantization (FedLQ [37]) and multi-grained dequantization strategies (FedMQ+ [38]), achieved improved trade-offs between accuracy and efficiency. Most existing techniques, however, treated sparsification, quantization, and encoding as separate processes, with limited emphasis on end-to-end reconstruction fidelity. Addressing this gap, FedWSOComp [20] proposed a compression-aware framework that integrated deterministic k-means quantization, soft Top-z sparsification [40], and Huffman encoding. This approach enabled high-fidelity reconstruction while achieving compression levels of up to $\times 1085$, resulting in 99% savings in both bandwidth and energy during communication. Although FedWSOComp demonstrated promising results in brain tumour segmentation, its evaluation primarily focused on scalability with increasing client numbers and fixed clustering and sparsification parameters.

Building on these findings, the present study investigated performance stability under varying heterogeneity, clustering, and sparsification conditions, thereby extending the applicability and usability of the original framework.

3. Materials and methods

3.1. Materials

This study employed the public and multi-institutional FeTS 2024 dataset [41]. This challenge extended the BraTS framework [42] in a federated setting, where images are located at local institutions. This condition is often observed in hospital networks, where privacy and legal concerns pose challenges to the sharing of raw imaging data. Specifically, it integrates clinically acquired scans originating from the RSNA-ASNR-MICCAI BraTS 2021 challenge, the BraTS 2023 Glioma challenge, as well as additional independent datasets contributed by remote institutions participating in a real-world federated network [41]. As a result, the dataset encompasses substantial variability in acquisition protocols, scanners, patient populations, tumour appearance, and institutional data distributions.

The dataset includes multi-parametric MRI (mpMRI) images of the brain of 1251 different patients, each with a corresponding expert-labelled segmentation mask defining tumour subregions. Each scan consists of 4 modalities: i. T1-weighted, ii. T1-contrast enhanced (T1ce), iii. T2-weighted and iv. T2-FLAIR.

Although the dataset was collected from 33 real institutions, the metadata includes an artificial division into 23 partitions. In this study, the original partitions were used to create 4 major clients (or silos) with more data. This approach reduced the number of clients that needed to be trained while still enabling the simulation of non-uniform local datasets, which are characteristic of clinical FL implementations.

Using these 4 major clients, both homogeneous and heterogeneous partitioning arrangements have been defined. In the IID partitioning, 100 random samples per client were selected from each client. This ensured that each client received an equal number of scans and a roughly similar distribution of tumour characteristics. This scenario represented an optimistic and ideal case of FL in which clients work in balanced and similar data environments and served as a baseline for further analysis. For non-IID partitioning, the major clients were kept unbalanced. This strategy simulated a more challenging and more representative scenario of real-world clinical conditions, in which institutions differed significantly in both the quantity of available patient records and the quality of imaging. For each local client, the 10% of the total cases were used for the local validation set, 10% for the local testing set and remaining 80% for the local training set. Nevertheless, to ensure a clearer evaluation protocol, an independent testing set of 67 patients was used, differing from the original FedWSOComp study. This separation between training and testing data reflected a more realistic deployment setting and aimed to reduce potential evaluation bias. A comprehensive overview of the partitioning strategies employed is provided in Table 1. In addition, data distribution among clients was presented in Fig. 1.

Table 1
Client-level data allocation for IID and Non-IID configurations, with a separate independent test set on the server.

IID			
Client	# Patients	Percentage	Original partitions
Client 1	100	25%	1
Client 2	100	25%	18
Client 3	100	25%	4, 6, 13
Client 4	100	25%	2, 3, 5, 7, 8, 9, 10, 11, 12
Total	**400**	**100.0%**	
Non-IID			
Client	# Patients	Percentage	Original partitions
Client 1	511	43%	1
Client 2	382	32%	18
Client 3	151	13%	4, 6, 13, 21
Client 4	140	12%	2, 3, 5, 7, 8, 10, 11, 12, 14, 15, 17, 19, 22, 23
Total	**1184**	**100.0%**	
Independent test set			
	# Patients	Percentage	Original partitions
Server	**67**	**100%**	9, 16, 20

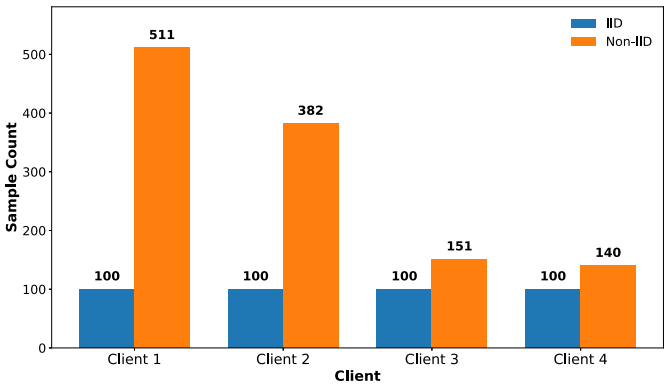


Fig. 1. The number of samples per client both in IID (blue) and non-IID (orange) settings. In an IID settings, clients received an equal number of patient cases. In a non-IID setting, the assignment of mixed data partitions to clients resulted in unbalanced local datasets, thereby simulating a real-world scenario.

3.2. Methods

3.2.1. Data preprocessing

The MONAI [43] framework was employed to implement the medical segmentation workflow, including the data preprocessing pipeline. To standardise spatial orientation across clients, all images were reoriented to the canonical RAS (Right-Anterior-Superior) coordinate system. Furthermore, images and labels were resampled to an isotropic voxel spacing of $1.0 \times 1.0 \times 1.0$ mm^3, ensuring uniform spatial resolution across the dataset. Constant padding was applied to match the input dimensions to $256 \times 256 \times 160$ voxels. A min-max intensity normalization was applied for each modality, scaling values to the $[0, 1]$ range. During training, random data augmentation was applied in a deterministic manner at the client level, including flipping along the spatial axes, intensity scaling, and intensity shifting.

3.2.2. Deep learning architecture

While the original FedWSOcomp study [20] employed a 2D U-Net [5], we adopted a 3D U-Net to better exploit volumetric information in MRI data [44]. As discussed in Section 2, this architecture has been widely used in biomedical image segmentation tasks due to its encoder–decoder structure and skip connections, which preserve spatial detail by combining high-resolution features from the contracting path with those from the expanding path. The implemented model consisted of five hierarchical levels. At each downsampling stage, the number of feature channels was doubled, starting from 16 up to 256. Each block applied a $3 \times 3 \times 3$ convolution, followed by batch normalization and ReLU activation. Downsampling was performed via max pooling with a stride of 2. In the decoding path, transposed convolutions were used to upsample the feature maps, which were concatenated with the corresponding encoder features via skip connections to preserve localisation accuracy. A dropout rate of 0.2 was applied at the bottleneck to mitigate overfitting. The input to the model was a four-channel 3D volume comprising the T1, T1ce, T2, and FLAIR MRI modalities. A single-channel output volume was obtained through sigmoid activation, representing voxel-wise probabilities of tumour presence. The model was trained using Dice loss, which directly measured the overlap between predicted tumour masks and ground truth annotations.

3.2.3. Aggregation strategy

The FedWSOComp [20] strategy was built upon the widely adopted FedAvg algorithm [8], introducing quantization-aware modifications to reduce communication overhead between clients and the central server. While standard FedAvg involved the transmission of full-precision model updates, FedWSOComp employed compression techniques to enable scalability across numerous clients, particularly within bandwidth-constrained settings such as federated medical image segmentation [20]. The training process began with the initialisation of the global model weights $w^{(0)}$ by the server, which then broadcast them to all participating clients. Each client trained the model locally over a fixed number of epochs, producing an updated model $w_i^{(t)}$ and computing the local update as:

$$\Delta w_i^{(t)} = w_i^{(t)} - w^{(t)} \tag{1}$$

Rather than transmitting $\Delta w_i^{(t)}$ directly, clients compressed their updates into binary representations $C_i^{(t)}$ using the FedWSOComp encoding pipeline. Only the compressed updates were transmitted to the server. Upon receiving the encoded updates, the server executed the following operations:

- Huffman decoding to recover the quantised index sequence z_i;
- Index-to-centroid mapping to reconstruct the quantised update:

$$z_i \rightarrow \hat{w}_i = \mu_{z_i} \tag{2}$$

- Aggregation of the reconstructed updates via a weighted FedAvg scheme.

The global model was subsequently updated as:

$$w^{(t+1)} = w^{(t)} + \sum_{i=1}^{N} \frac{n_i}{\sum_j n_j} \cdot \Delta \tilde{w}_i \tag{3}$$

where n_i denoted the number of local samples held by client i, and $\Delta \tilde{w}_i$ represented the decoded and reconstructed update provided by that client. Following aggregation, the updated global model $w^{(t+1)}$ was redistributed to all clients to initiate the next training round. This process was repeated for a predefined number of communication rounds. All compression and decompression steps were deterministic, ensuring round-trip consistency without the need for additional metadata. This design led to considerable communication savings, as also reported in Onaizah et al. [20]. As in the original FedWSOComp algorithm, compression was performed independently for each layer, treating model parameters as separate tensors. Sparsification was applied using a top-z scheme [45], in which each tensor was pruned by identifying the z-quantile of absolute magnitudes and clipping smaller values while preserving their sign. This approach reshaped the update distribution without introducing zero values and was shown to improve convergence, particularly for bias tensors when using low sparsification thresholds [20]. Following sparsification, each tensor was quantised via deterministic k-means clustering, with centroids computed per layer. Unlike stochastic or fixed-level quantization schemes, this method considered the statistical distribution of tensor values, thereby reducing update variance and preserving representativeness. Raw values were replaced by cluster indices, which were subsequently encoded using entropy-based methods such as Huffman coding. More frequent indices were assigned shorter bit-lengths, further enhancing compression efficiency [20]. An overview of the complete FedWSOComp pipeline employed in this study, which illustrates the full compression workflow, is provided in Fig. 2.

3.2.4. Compression parameter analysis

The FedWSOComp strategy relies on two principal compression parameters: the number of quantization clusters (C) and the sparsification rate (S_R). In the original study [20], the number of clusters was fixed at $C = 3$, and sparsification was applied conservatively. The authors reported improved convergence, particularly for bias tensors, when sparsification thresholds were set below the Top-0.2–thus 20%–level. In the present study, we extended this analysis by systematically exploring a broader range of compression configurations. Specifically, the number of clients was held constant while performing a grid search over the two parameters, as follows:

- Number of clusters $C \in \{3, 32, 64\}$

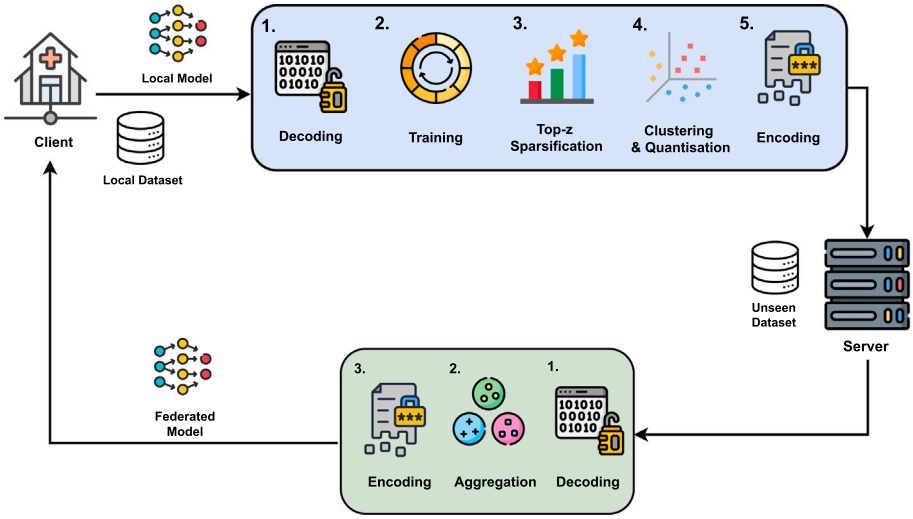


Fig. 2. Overview of the complete FedWSOComp pipeline adopted in this study.

- Soft Top-z sparsification rate $S_R \in \{20, 30, 60\}$

All possible combinations were evaluated in both IID and non-IID scenarios, yielding 18 distinct configurations. Each setting was tested using the complete FedWSOComp compression and aggregation pipeline, allowing for an isolated assessment of the relationship between compression intensity and segmentation performance under otherwise identical experimental conditions.

3.2.5. Ablation study

To disentangle the individual contributions of the different components of the FedWSOComp framework, we performed a structured ablation study under both IID and non-IID data distributions. The objective was to isolate the functional impact of each mechanism, by selectively disabling or activating specific components while keeping all other experimental conditions unchanged. Therefore, three additional experimental pipelines were implemented:

1. *FedAvg baseline.* The FedAvg strategy was used without any compression, clustering, or sparsification mechanisms. This configuration served as a reference upper-bound baseline in terms of communication fidelity and segmentation performance, allowing us to quantify the effect introduced by FedWSOComp.
2. *Clustering-only configuration (C-only).* In this setting, only the clustering-based quantization component of FedWSOComp was enabled. The number of quantization clusters was varied as $C \in \{3, 32, 64\}$, while the sparsification rate was fixed to $S_R = 0$, effectively disabling gradient sparsification.
3. *Sparsification-only configuration (S-only).* With this configuration, the clustering mechanism was disabled and only the Soft Top-z sparsification component was applied. The sparsification rate was varied as $S_R \in \{20\%, 30\%, 60\%\}$, allowing an isolated analysis of the impact of gradient pruning on model convergence and segmentation performance in the absence of quantization.

All ablation configurations were evaluated under both IID and non-IID client distributions, using the same number of clients, model architecture, training schedule, and data partitions adopted in the full FedWSOComp experiments. This controlled design ensured that any observed performance variations could be attributed exclusively to the inclusion or removal of specific FedWSOComp components, enabling a principled decomposition of their individual and combined contributions to communication efficiency and learning performance.

3.2.6. Communication efficiency evaluation

To enable a principled and reproducible analysis of communication efficiency across different the configurations, we defined a unified evaluation protocol for measuring transmission costs and compression effectiveness. The objective was to quantify the impact of each compression strategy on the communication payload exchanged between clients and server, independently of model performance metrics. For each experimental configuration presented in Section 3.2.5, the communication cost was computed as the total number of transmitted bits per communication round, accounting exclusively for model updates exchanged during the aggregation phase.

The baseline transmission cost was defined using the uncompressed FedAvg protocol, where full-precision model parameters are transmitted without any form of compression, quantization, or sparsification.

This value provided a fixed reference point for all subsequent comparisons.

Communication efficiency was quantified using two complementary metrics:

- the *compression ratio*, defined as the ratio between the FedAvg baseline payload and the compressed payload size, providing a multiplicative measure of communication reduction.
- the *payload reduction percentage*, defined as the relative decrease in transmitted bits with respect to the FedAvg baseline.

These metrics allowed for a consistent comparison across compression mechanisms, including the single components ablation study, and their combined application. All communication measurements were performed under identical experimental conditions, including model architecture, number of clients, training schedule, aggregation protocol, and data distribution, ensuring that differences in transmission cost arose solely from the adopted compression strategy.

3.2.7. Experimental DL protocol and metrics

The federation was implemented using the Flower framework [46], while the DL components were developed using PyTorch [47] and MONAI [2] [43]. Model optimization was performed using the Adam optimiser, with a learning rate set to 1×10^{-4} and a weight decay of 1×10^{-5}. A cosine annealing learning rate scheduler was employed, with the cycle

[2] All experiments were conducted on a Supermicro AS-4124GS-TNR 4U Rackmount GPU SuperServer equipped with an NVIDIA A100 80 GB GPU, 16 CPU cores, and 128 GB of RAM

length set to the number of local training epochs. Mixed-precision training was enabled via automatic loss scaling using PyTorch *GradScaler*. Due to the substantial memory requirements of 3D training, the batch size was set to 1. For all experiments, clients were trained for 25 rounds, each consisting of 5 local epochs.

Segmentation performance was primarily assessed using standard medical imaging segmentation metrics:

- the Dice Similarity Coefficient (DSC), which quantified the volumetric overlap between the predicted and ground truth tumour masks. The DSC, bounded between 0 and 1, was defined as:

$$DSC = \frac{2|A \cap B|}{|A| + |B|} \tag{4}$$

  where A and B denoted the predicted and ground truth binary masks, respectively.
- the 95th percentile Hausdorff Distance (HD95) [48] was computed to evaluate spatial agreement between segmentations, offering complementary insight into anatomical accuracy. The HD95 metric corresponded to the 95th percentile of all pairwise distances, reducing the influence of outliers.

4. Results

4.1. Compression parameter analysis

As reported in Section 3.2.4, the FedWSOComp strategy was assessed under 18 different compression configurations, by varying both the quantization cluster size and the Soft Top-z sparsification retention ratio, in the context of IID and non-IID data distributions. Tables 2 and 3 summarise the performance of the best-performing models, defined as those achieving the highest validation performance during training.

These models were subsequently evaluated on an independent test dataset that remained unseen during training. Both the global server models and the local clients models were considered, allowing the assessment of cross-site generalization as well as client-specific adaptation.

For Client 1, under non-IID conditions with $C = 64$ and $S_R = 60\%$, the local model achieved a DSC of 0.99 ± 0.03 and an HD95 of 5.15 ± 4.17 mm, indicating high segmentation quality when compression is lower. The federated model, trained under the same parameters and evaluated on the independent test set, obtained a DSC of 0.98 ± 0.09 and an HD95 of 10.40 ± 15.54 mm, confirming that the applied sparsification ratio retained sufficient information to preserve boundary fidelity in both global and local models. In IID conditions, DSC values peaked at $S_R = 30\%$ with $C = 32$ (0.85 ± 0.24), although further increasing S_R to 60% improved boundary accuracy, yielding an HD95 of 9.75 ± 7.38 mm at $C = 64$, with a slight trade-off in DSC (0.81 ± 0.31). This suggested that, in IID settings, volumetric overlap was maximised at $S_R = 30\%$, while boundary accuracy improved at higher retention. Interestingly, this trade-off disappeared under non-IID conditions, where high C values combined with $S_R = 60\%$ yielded the best overall performance across both metrics.

For Client 2, which displayed greater intra-dataset variability, the best non-IID outcome was reached at $C = 64$, $S_R = 60\%$, with DSC 0.79 ± 0.34 and HD95 10.19 ± 8.24 mm. Here, boundary accuracy surpassed that of the global model, while DSC remained comparable. In IID settings, results were more consistent across parameter combinations. The best outcome was again observed with $C = 64$, $S_R = 60\%$ (DSC 0.53 ± 0.44, HD95 11.64 ± 8.37 mm), with a marginal DSC improvement at $C = 32$, $S_R = 60\%$ (DSC 0.56 ± 0.47, HD95 12.87 ± 9.14 mm).

For Client 3, the non-IID setting with $C = 32$, $S_R = 60\%$ resulted in a local HD95 of 11.95 ± 9.24 mm. The corresponding global model under the same parameters achieved DSC 0.98 ± 0.12 and HD95 13.74 ± 15.71 mm, reflecting stable performance. In IID conditions, the lowest HD95 was observed at $C = 32$, $S_R = 60\%$ (11.95 ± 9.24 mm), whereas the highest DSC was obtained at $C = 64$, $S_R = 20\%$ (0.92 ± 0.13), though at the expense of substantially higher HD95.

For Client 4, segmentation accuracy reached its maximum, with non-IID DSC values of 1.00 ± 0.00 in several configurations. However, boundary precision was not always preserved, as illustrated by $C = 64$, $S_R = 60\%$ (non-IID), which achieved HD95 of 10.40 ± 15.54 mm despite near-perfect DSC. In IID conditions, DSC peaked at 0.99 ± 0.04 with $C = 32$, $S_R = 30\%$ and $C = 64$, $S_R = 30\%$, while boundary accuracy further improved at higher retention (e.g., $C = 64$, $S_R = 60\%$, HD95 14.18 ± 9.96 mm).

A comparison of the results obtained on the independent set (reported as "Server" in Tables 2 and 3) revealed a consistent pattern: moderate-to-fine quantization ($C = 32$ or $C = 64$) combined with higher retention ($S_R = 60\%$) yielded the most favourable balance between DSC and HD95. Specifically, under IID, $C = 32$ or 64 with $S_R = 60\%$ achieved DSC values of 0.98 ± 0.12 and 0.98 ± 0.13, and HD95 of 13.74 ± 15.71 mm and 14.11 ± 15.74 mm, respectively. Under non-IID conditions, the same retention ratio with $C = 64$ produced optimal performance (DSC 0.98 ± 0.09, HD95 10.40 ± 15.54 mm). Conversely, low retention ($S_R = 20\%$)

Table 2
Test results under IID data distribution for different combinations of C and S_R. Dice Similarity Coefficient (DSC) and 95th percentile Hausdorff Distance (HD95) are reported as mean ± standard deviation for the server and each client.

C	S_R (%)	Metric	Server	Client 1	Client 2	Client 3	Client 4
3	20	DSC	0.77 ± 0.21	0.24 ± 0.12	0.21 ± 0.17	0.69 ± 0.32	0.74 ± 0.24
		HD95	69.23 ± 13.28	81.00 ± 17.40	88.44 ± 7.26	69.28 ± 16.99	68.60 ± 11.51
3	30	DSC	0.93 ± 0.14	0.47 ± 0.25	0.40 ± 0.40	0.78 ± 0.26	0.94 ± 0.13
		HD95	57.70 ± 19.48	74.02 ± 19.25	85.46 ± 9.93	59.33 ± 24.69	57.16 ± 16.56
3	60	DSC	0.90 ± 0.14	0.63 ± 0.28	0.37 ± 0.34	0.71 ± 0.26	0.93 ± 0.09
		HD95	60.47 ± 18.88	75.41 ± 18.48	86.90 ± 10.44	58.08 ± 22.77	63.19 ± 15.99
32	20	DSC	0.95 ± 0.13	0.54 ± 0.25	0.46 ± 0.37	0.89 ± 0.18	0.94 ± 0.08
		HD95	61.23 ± 18.73	73.35 ± 19.42	83.39 ± 12.31	60.08 ± 21.94	61.71 ± 16.63
32	30	DSC	0.97 ± 0.12	0.85 ± 0.24	0.43 ± 0.40	0.78 ± 0.33	0.99 ± 0.04
		HD95	25.25 ± 24.93	26.95 ± 20.46	46.95 ± 35.99	24.04 ± 13.53	22.24 ± 18.19
32	**60**	**DSC**	**0.98 ± 0.12**	**0.80 ± 0.32**	**0.56 ± 0.47**	**0.75 ± 0.40**	**0.98 ± 0.07**
		HD95	**13.74 ± 15.71**	**10.69 ± 6.95**	**12.87 ± 9.14**	**11.95 ± 9.24**	**14.20 ± 9.84**
64	20	DSC	0.95 ± 0.12	0.56 ± 0.27	0.48 ± 0.37	0.92 ± 0.13	0.97 ± 0.05
		HD95	58.60 ± 20.52	72.47 ± 19.76	80.90 ± 14.01	58.31 ± 22.91	59.71 ± 18.02
64	30	DSC	0.97 ± 0.12	0.85 ± 0.24	0.45 ± 0.39	0.80 ± 0.31	0.99 ± 0.04
		HD95	26.59 ± 25.50	32.42 ± 19.03	48.32 ± 36.29	25.41 ± 16.46	27.74 ± 21.01
64	60	DSC	0.98 ± 0.13	0.81 ± 0.31	0.53 ± 0.44	0.75 ± 0.40	0.97 ± 0.08
		HD95	14.11 ± 15.74	9.75 ± 7.38	11.64 ± 8.37	12.23 ± 8.87	14.18 ± 9.96

Table 3
Test results under non-IID data distribution for different combinations of C and S_R. Dice Similarity Coefficient (DSC) and 95th percentile Hausdorff Distance (HD95) are reported as mean ± standard deviation for the server and each client.

C	S_R (%)	Metric	Server	Client 1	Client 2	Client 3	Client 4
3	20	DSC	0.92 ± 0.15	0.55 ± 0.28	0.42 ± 0.38	0.82 ± 0.23	0.92 ± 0.14
		HD95	61.35 ± 18.48	74.32 ± 18.77	82.08 ± 13.39	58.33 ± 22.56	61.08 ± 15.35
3	30	DSC	0.96 ± 0.11	0.76 ± 0.28	0.45 ± 0.38	0.83 ± 0.23	0.99 ± 0.03
		HD95	54.02 ± 22.78	69.89 ± 19.51	71.85 ± 25.79	41.03 ± 29.92	51.61 ± 22.15
3	60	DSC	0.98 ± 0.09	0.83 ± 0.23	0.48 ± 0.37	0.84 ± 0.22	1.00 ± 0.00
		HD95	50.21 ± 25.27	68.50 ± 16.99	69.82 ± 27.20	39.93 ± 27.89	47.92 ± 26.84
32	20	DSC	0.98 ± 0.12	0.83 ± 0.25	0.42 ± 0.40	0.81 ± 0.30	1.00 ± 0.00
		HD95	37.66 ± 29.13	50.16 ± 10.80	69.19 ± 19.19	35.92 ± 29.82	38.13 ± 26.86
32	30	DSC	0.98 ± 0.13	0.85 ± 0.25	0.40 ± 0.41	0.77 ± 0.36	0.99 ± 0.04
		HD95	27.02 ± 27.49	32.43 ± 21.42	49.31 ± 33.66	33.27 ± 27.47	28.79 ± 26.55
32	60	DSC	0.97 ± 0.14	0.70 ± 0.44	0.31 ± 0.45	0.70 ± 0.46	0.91 ± 0.26
		HD95	13.34 ± 15.23	87.39 ± 68.91	17.93 ± 12.65	15.21 ± 19.05	12.10 ± 7.62
64	20	DSC	0.98 ± 0.12	0.82 ± 0.25	0.48 ± 0.37	0.84 ± 0.25	1.00 ± 0.00
		HD95	44.74 ± 27.34	59.76 ± 9.61	74.73 ± 15.02	37.32 ± 29.96	45.79 ± 23.61
64	30	DSC	0.97 ± 0.14	0.90 ± 0.21	0.37 ± 0.42	0.74 ± 0.40	0.95 ± 0.11
		HD95	22.25 ± 24.93	29.58 ± 21.86	43.54 ± 34.15	28.92 ± 23.14	28.23 ± 27.08
64	**60**	**DSC**	**0.98 ± 0.09**	**0.99 ±0.03**	**0.79 ± 0.34**	**0.83 ± 0.34**	**0.98 ± 0.05**
		HD95	**10.40 ± 15.54**	**5.15 ± 4.17**	**10.19 ± 8.24**	**11.79 ± 10.45**	**17.56 ± 15.31**

combined with coarse quantization ($C = 3$) proved least effective, with IID DSC as low as 0.77 ± 0.21 and HD95 69.23 ± 13.28 mm, and non-IID DSC 0.92 ± 0.15 with HD95 61.35 ± 18.48 mm.

Fig. 3 further illustrates these trends. In IID settings, $C = 32$ with $S_R = 60\%$ produced smooth, monotonic DSC improvements with minimal oscillations, indicative of stable convergence. In contrast, coarse quantization with low retention (e.g., $C = 3$, $S_R = 20\%$) resulted in unstable trajectories, with rapid degradation of DSC during training. In non-IID conditions, configurations with high retention ($S_R = 60\%$) and fine quantization ($C = 64$) retained DSC above 0.95 throughout, demonstrating robustness against heterogeneous updates. In addition, to complement the quantitative findings, a fixed-slice comparison was performed to highlight qualitative differences between configurations, as shown in Figs. 4 and 5 for IID and non-IID settings, respectively.

4.2. Ablation study

As described in Section 3.2.5, to assess the individual contribution of each component of the *FedWSOcomp* framework, we conducted an ablation study under both IID and non-IID data distributions. In the following Section, we report the results of each ablation configuration, including a systematic comparison with the full FedWSOcomp pipeline using the reference configuration $C = 32$ and $S_R = 60\%$.

The results reported in Tables 4 and 5 correspond to the best-performing communication rounds, thus the models achieving the highest validation performance during the federated training process. Performance is reported for both the server (independent test set) and the clients.

Under IID conditions (Table 4), the FedAvg baseline achieved strong segmentation performance at both server and client levels, with a noticeable degradation on Client 2, which –as also reported in Section 4.1– exhibited the highest intra-dataset variability.

When quantization was isolated ($C - only$), with coarse quantization ($C = 3$), segmentation fidelity deteriorated substantially across both metrics, indicating that the transmitted updates were excessively distorted to preserve tumour boundaries. In contrast, moderate-to-fine quantization ($C = 32, C = 64$) preserved both overlap and boundary accuracy at the server level, achieving an average server-side performance of 0.98 ± 0.12 and HD95 10.82 ± 15.72 mm for $C = 32$, and DSC 0.99 ± 0.02 with HD95 9.01 ± 11.62 mm for $C = 64$. In the latter configuration, HD95 improved consistently both at the server and client levels, indicating enhanced boundary preservation. This behaviour can be attributed to the sensitivity of boundary-based metrics to inconsistencies in local gradient directions and anatomical representations across heterogeneous clients. While FedAvg performs a global parameter averaging that can smooth and distort fine-grained boundary features, clustering-based aggregation groups clients with similar data distributions, thereby preserving coherent geometric and anatomical structures.

When sparsification was evaluated independently ($S_R - only$), with low-to-moderate sparsification ($S_R = 20\%, S_R = 40\%$), segmentation quality was already substantially degraded, as reflected by moderate DSC values and high HD95, indicating pronounced boundary distortions. With stronger sparsification ($S_R = 60\%$), volumetric overlap remained relatively high, with an average server-side DSC of 0.94 ± 0.13. However, boundary accuracy still deteriorated, with an average server-side HD95 of 41.05 ± 30.26 mm. These findings suggested that sparsification alone can preserve volumetric overlap while introducing significant boundary errors.

Under non-IID conditions (Table 5), the same overall trends were observed. FedAvg remained robust at both server and client levels. The $C - only$ configurations, showed that coarse clustering induced severe boundary degradation, with an average server-side HD95 of 68.69 ± 14.58 mm for $C = 3$, whereas finer configurations ($C = 32, C = 64$) preserved both overlap and boundary accuracy.

For the $S_R - only$ configurations, boundary degradation was consistently observed across sparsity ratios. At low sparsification ($S_R = 20\%$), although overlap remained relatively high, boundary accuracy was already severely compromised, indicating once again significant contour deviations.

The same behaviour was observed with moderate and higher sparsification ($S_R = 60\%$). Overlap remained high (i.e., server-side DSC of 0.98 ± 0.08), but boundary accuracy remained unsatisfactory compared to FedAvg and $C - only$ aggregation with $C \in \{32, 64\}$.

4.3. Communication efficiency evaluation

Communication efficiency results are reported in Table 6. The baseline transmission cost was constant across all modes (1.54×10^8 bits). The $C - only$ configurations achieved compression ratios ranging from ×3.13 to ×3.77, corresponding to payload reductions of 68.11%–73.48%. The highest compression was obtained for the coarsest clustering setting ($C = 3$, ×3.77), which, however, corresponds to the configuration

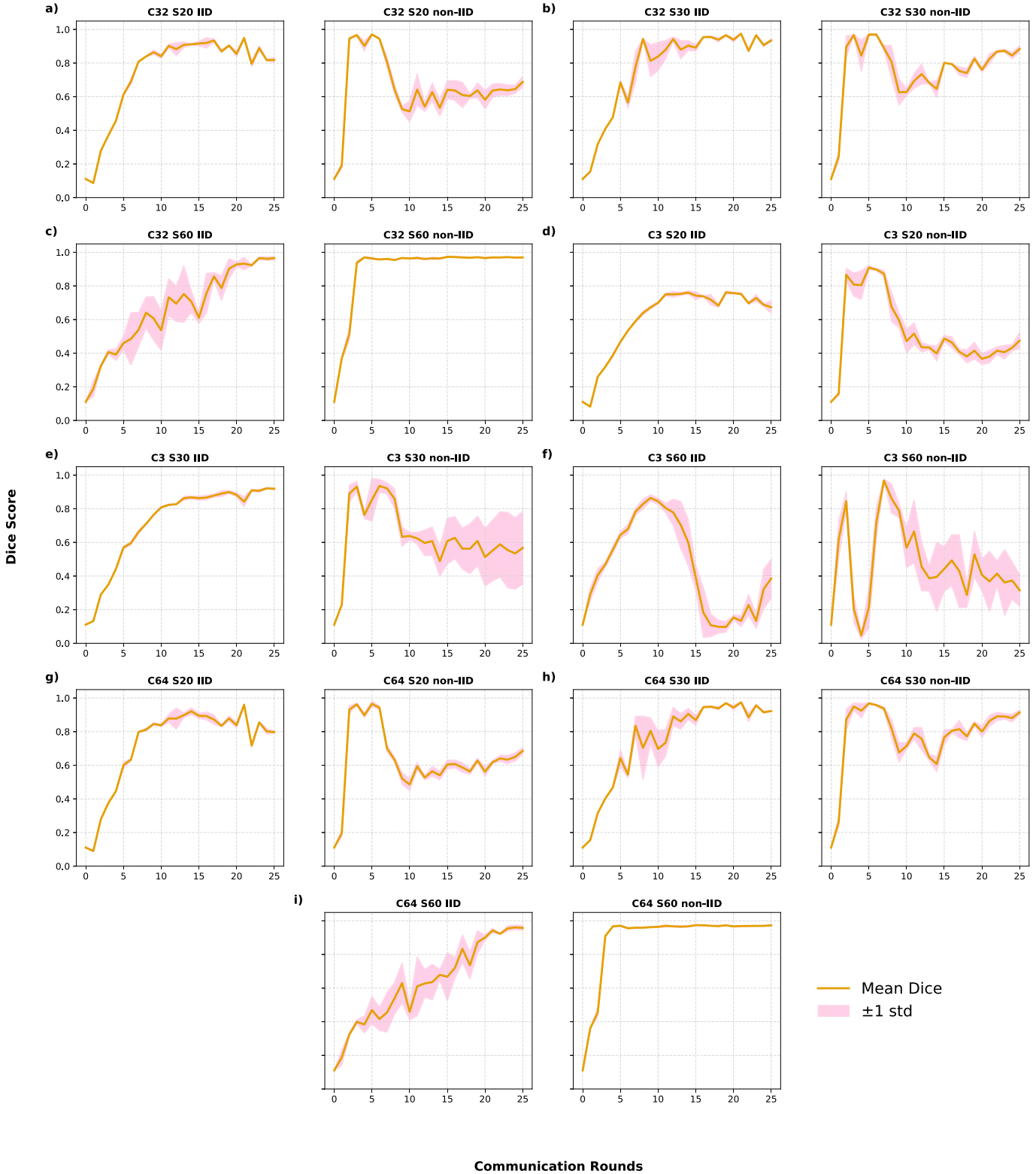


Fig. 3. Performance trends for varying S_R and C values under IID and non-IID settings.

yielding the poorest segmentation performance in Table 4, highlighting a clear trade-off between communication efficiency and model accuracy.

For the $S_R - only$ configurations, communication gains were strongly dependent on the sparsity level. Specifically, $S_R - only$ with $S_R = 20\%$ resulted in no effective reduction, matching the FedAvg baseline, while $S_R = 30\%$ produced only a negligible improvement (×1.01, 0.93%). Increasing sparsity to $S_R = 60\%$ led to a modest communication reduction (×1.24, 19.48%), which remained substantially weaker than the gains achieved by $C - only$ strategies.

In contrast, the combined configuration –reported as FedWSOcomp– using the $C = 32$ and $S_R = 60\%$ configuration achieved a good compression ratio of ×3.40 with a 70.65% payload reduction, providing communication savings comparable to $C - only$ while simultaneously preserving segmentation performance.

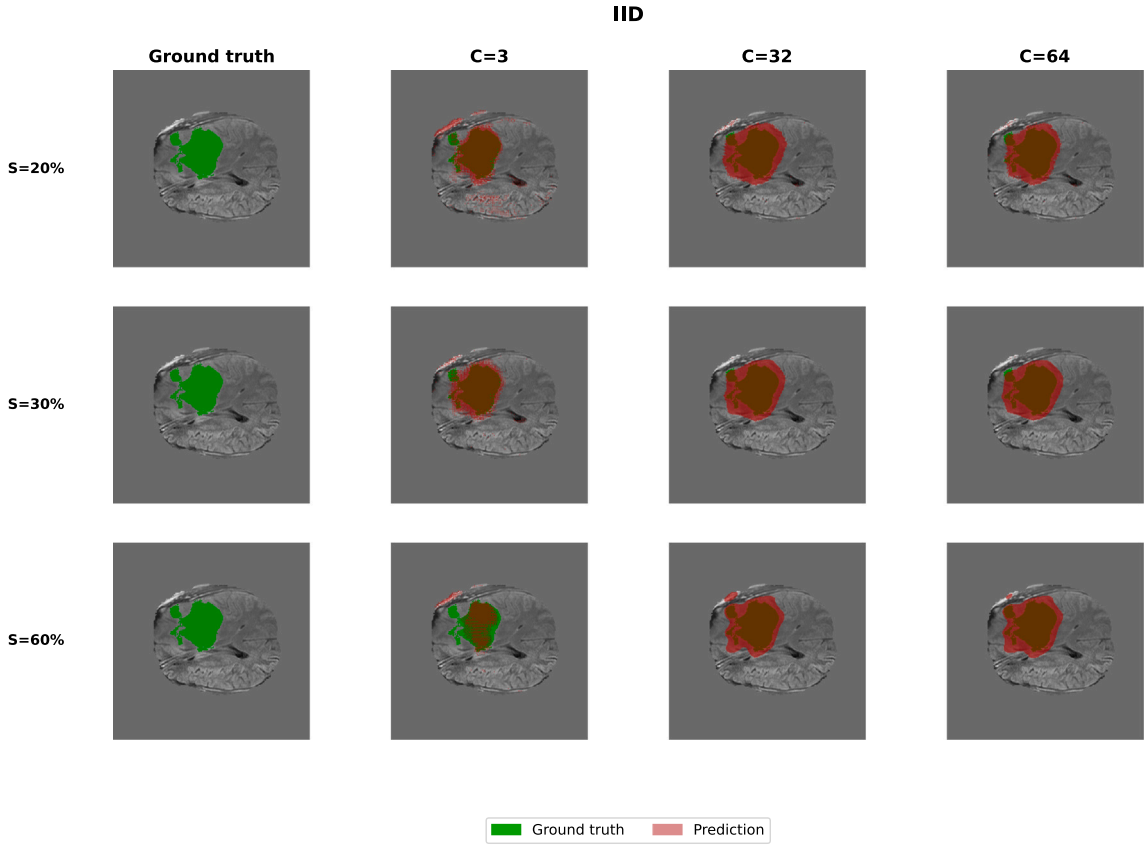


Fig. 4. Qualitative comparison of segmentation predictions across all compression configurations under IID conditions. Results are shown for different sparsification rates ($S_R = 20\%, 30\%, 60\%$) and quantization levels ($C = 3, 32, 64$), with the corresponding ground truth masks.

5. Discussion

The experimental evaluation of FedWSOComp under both IID (Table 2) and non-IID (Table 3) conditions provided insights into the relationship between sparsification, quantization, and federated aggregation in 3D brain tumour segmentation. At the global level, the results indicated that extreme compression ($C = 3$, $S_R = 20\%$) led to severe performance degradation, with the DSC falling below 0.80 and the HD95 exceeding 60 mm. These findings confirmed that aggressive sparsification combined with coarse quantization caused substantial information loss, limiting the global model's ability to capture tumour morphology. In contrast, configurations with higher retention ($S_R = 60\%$) and moderate-to-fine quantization ($C = 32$ or $C = 64$) consistently achieved the best trade-off between accuracy and communication efficiency. In non-IID settings, the combination of $C = 64$ and $S_R = 60\%$ reached better overall performance, maintaining high segmentation quality despite heterogeneous client distributions. This observation aligned with prior studies that emphasised the importance of retaining high-magnitude parameters to preserve representational capacity while reducing communication costs [16,40,49]. Conceptually, FedWSOComp shared similarities with the dynamic layer-wise sparsification strategy introduced by Zhang et al. [40], but it differed in its adoption of a fixed *soft* Top-z policy. This softer retention mechanism appeared effective in preserving fine anatomical structures, which are particularly relevant in volumetric tumour segmentation.

In addition to the quantitative analysis, the qualitative results presented in Figs. 4 and 5 provide further insight into the behaviour of FedWSOComp under different compression settings. In the IID case, coarse quantization with $C = 3$ and low retention ($S_R = 20\%$) produced poorly defined segmentations, including spurious regions extending beyond the brain volume. Increasing retention to $S_R = 30\%$ and $S_R = 60\%$ reduced these artefacts, but the segmentations remained poor. Using finer quantization ($C = 32$ and $C = 64$), the predictions covered the ground truth more consistently, with only minor deviations along the outer lesion boundaries. However, the difference between $C = 32$ and $C = 64$ was negligible, indicating limited benefit from doubling the number of clusters. Under non-IID conditions, the larger number of samples available to each client led to more precise segmentations overall. As in the IID setting, $C = 3$ resulted in inaccurate predictions, particularly at $S_R = 20\%$ and $S_R = 30\%$. An improvement was noticed only using $S_R = 60\%$, although the model still failed to match the ground truth with high accuracy. In contrast, with $C = 32$ and $C = 64$, segmentation quality clearly improved as retention increased, yielding satisfactory results at $S_R = 30\%$ and excellent delineations at $S_R = 60\%$. Overall, these qualitative observations confirmed quantitative results. A coarse quantization ($C = 3$) consistently degraded segmentation quality, with the degradation being most pronounced when fewer training cases were available (IID scenario). By contrast, moderate-to-fine quantization ($C = 32$ or $C = 64$) combined with higher retention produced more accurate and reliable lesion delineations, particularly in heterogeneous non-IID scenarios.

Notably, substantial variability was observed across clients. For instance, Client 2, trained on a more heterogeneous dataset, exhibited reduced stability, whereas Client 4, with more homogeneous data, achieved near-perfect DSC. However, higher DSC values were sometimes associated with increased HD95, suggesting a trade-off between volumetric overlap and boundary precision. Such site-specific differences reinforced the need to assess FL models at both independent institutions and individual clients levels, in agreement with recent federated studies [10,11,34].

Under IID conditions, $C = 32$ provided the most favourable balance, achieving tighter variance bounds and higher DSC compared with $C = 3$ or $C = 64$. Although $C = 64$ yielded marginal robustness advantages

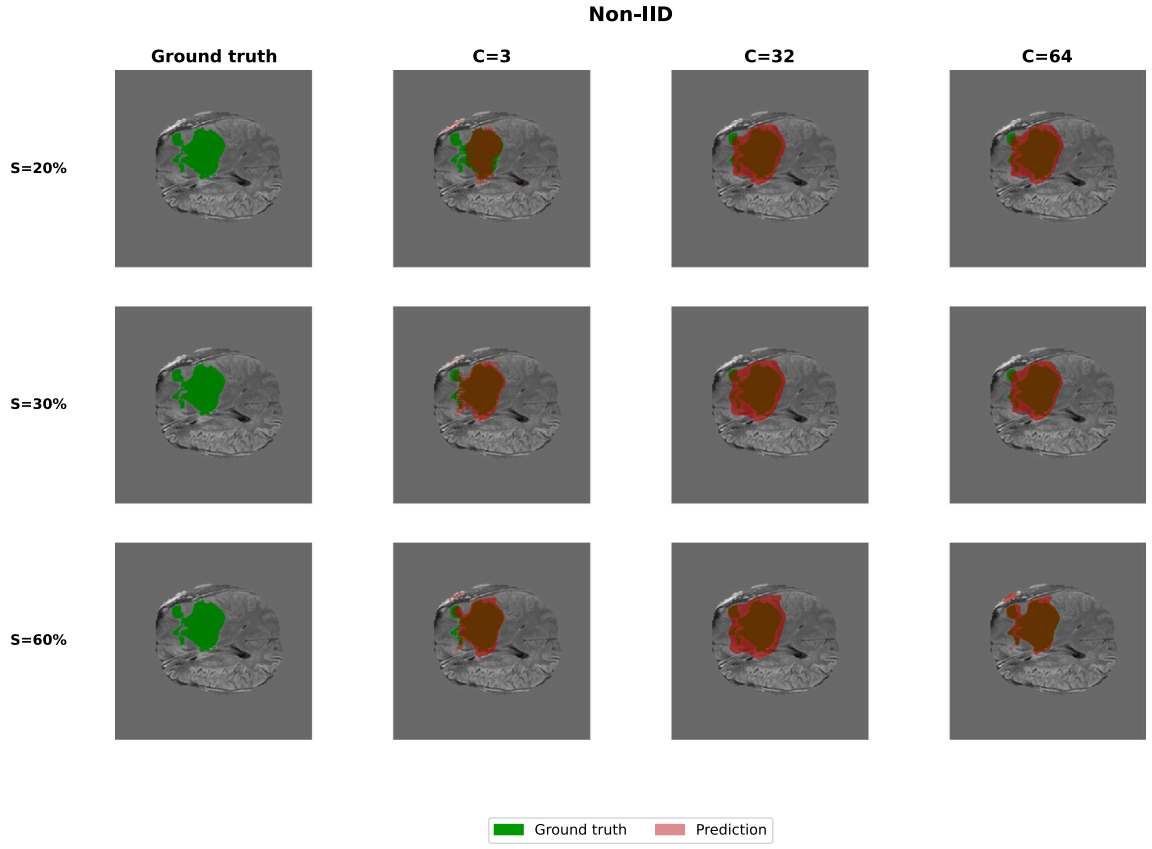


Fig. 5. Qualitative comparison of segmentation predictions across all compression configurations under non-IID conditions. Results are shown for different sparsification rates ($S_R = 20\%, 30\%, 60\%$) and quantization levels ($C = 3, 32, 64$), with the corresponding ground truth masks.

Table 4
Test results under IID data distribution for the four experiment modes: FedAvg, CH ($S_R = 0$, $C \in \{3, 32, 64\}$), SH ($C = 0$, $S_R = 60\%$), and SKH ($C = 32$, $S_R = 60\%$). Dice Similarity Coefficient (DSC) and 95th percentile Hausdorff Distance (HD95) are reported as mean ± standard deviation for the server and each client.

Pipeline	C	S_R (%)	Metric	Server	Client 1	Client 2	Client 3	Client 4
FedAvg	**0**	**0**	**DSC**	**0.99 ± 0.04**	**0.90 ± 0.30**	**0.71 ± 0.39**	**0.98 ± 0.05**	**1.00 ±0.00**
			HD95	**12.58 ± 18.87**	**12.32 ± 16.79**	**11.30 ± 9.18**	**11.39 ± 9.55**	**14.12 ± 13.54**
$C-only$	3	0	DSC	0.59 ± 0.21	0.46 ± 0.26	0.40 ± 0.28	0.54 ± 0.24	0.68 ± 0.18
			HD95	73.28 ± 12.69	76.19 ± 15.31	76.47 ± 14.14	73.21 ± 13.12	64.77 ± 8.36
$C-only$	32	0	DSC	0.98 ± 0.12	0.88 ± 0.30	0.63 ± 0.45	0.79 ± 0.34	1.00 ± 0.00
			HD95	10.82 ± 15.72	13.48 ± 18.47	12.50 ± 9.42	10.79 ± 9.57	14.04 ± 12.51
C − only	**64**	**0**	**DSC**	**0.99 ± 0.02**	**0.89 ± 0.30**	**0.78 ± 0.32**	**0.97 ± 0.07**	**1.00 ± 0.00**
			HD95	**9.01 ± 11.62**	**8.55 ± 7.97**	**9.07 ± 7.77**	**8.07 ± 7.30**	**13.37 ± 13.85**
S_R-only	0	20	DSC	0.74 ± 0.23	0.26 ± 0.14	0.40 ± 0.32	0.87 ± 0.17	0.64 ± 0.16
			HD95	61.49 ± 18.29	74.59 ± 17.84	83.43 ± 12.43	61.53 ± 21.26	60.95 ± 16.42
S_R-only	0	30	DSC	0.74 ± 0.17	0.30 ± 0.14	0.40 ± 0.27	0.70 ± 0.28	0.65 ± 0.27
			HD95	68.89 ± 12.76	77.19 ± 17.07	85.70 ± 9.29	68.51 ± 16.51	69.02 ± 11.55
S_R − only	**0**	**60**	**DSC**	**0.94 ± 0.13**	**0.72 ± 0.28**	**0.80 ± 0.31**	**1.00 ± 0.00**	**0.92 ± 0.08**
			HD95	**41.05 ± 30.26**	**30.61 ± 20.60**	**32.73 ± 31.09**	**27.14 ± 18.73**	**42.50 ± 28.49**

under non-IID distributions, this came at the cost of increased communication. Similar behaviour was observed in high-retention scenarios ($S_R = 60\%$), consistent with the findings of Jia et al. [50], who showed that inappropriate parameter choices could amplify inter-client variance. In this work, deterministic k-means quantization combined with Huffman coding appeared to mitigate such instability by constraining convergence trajectories.

When compared with FedSparse [49], which primarily relied on sparse gradient updates, FedWSOComp integrated quantization directly into the sparsification pipeline and employed entropy coding to enable exact weight reconstruction. This design minimised communication payloads while maintaining boundary integrity–an essential aspect in 3D segmentation, where even minor inaccuracies may compromise clinical reliability.

Interestingly, the ablation results revealed that HD95 degradation was more pronounced under IID conditions than in non-IID settings for the FedAvg baseline. This counterintuitive behaviour was likely due to the limited number of patients used to construct a perfectly balanced IID partition, which may reduce boundary generalization in individual experiments. In contrast, the inherent heterogeneity of the non-IID setting appears to provide more stable anatomical variability, supporting improved generalization in some cases.

Table 5
Test results under non-IID data distribution for the four experiment modes: FedAvg ($C = 0$, $S_R = 0$), CH ($S_R = 0$, $C \in \{3, 32, 64\}$), SH ($C = 0$, $S_R = 60\%$), and SKH ($C = 32$, $S_R = 60\%$). Dice Similarity Coefficient (DSC) and 95th percentile Hausdorff Distance (HD95) are reported as mean ± standard deviation for the server and each client.

Pipeline	C	S_R (%)	Metric	Server	Client 1	Client 2	Client 3	Client 4
FedAvg	**0**	**0**	**DSC**	**0.99 ± 0.05**	**0.96 ± 0.17**	**0.83 ± 0.34**	**0.94 ± 0.16**	**0.97 ± 0.07**
			HD95	**8.99 ± 13.75**	**5.49 ± 3.88**	**7.84 ± 8.55**	**8.50 ± 7.97**	**13.67 ± 14.40**
$C-only$	3	0	DSC	0.85 ± 0.18	0.74 ± 0.27	0.60 ± 0.34	0.79 ± 0.22	0.91 ± 0.12
			HD95	68.69 ± 14.58	71.69 ± 15.16	73.57 ± 14.53	67.25 ± 15.19	61.13 ± 11.09
$C-only$	32	0	DSC	0.98 ± 0.08	0.98 ± 0.14	0.88 ± 0.31	0.94 ± 0.24	0.97 ± 0.09
			HD95	9.91 ± 14.47	6.09 ± 5.29	7.13 ± 7.31	7.80 ± 7.21	15.39 ± 15.27
C – only	**64**	0	**DSC**	**0.99 ± 0.05**	**0.97 ± 0.16**	**0.84 ± 0.33**	**0.94 ± 0.19**	**0.98 ± 0.08**
			HD95	**8.51 ± 13.03**	**5.51 ± 4.22**	**8.27 ± 9.27**	**8.33 ± 7.95**	**12.18 ± 15.27**
S_R-only	0	20	DSC	0.89 ± 0.17	0.93 ± 0.17	0.86 ± 0.24	0.94 ± 0.10	0.89 ± 0.17
			HD95	57.43 ± 19.32	56.07 ± 20.65	62.66 ± 16.77	48.63 ± 22.21	54.40 ± 13.41
S_R-only	0	30	DSC	0.99 ± 0.05	0.96 ± 0.12	0.88 ± 0.23	0.97 ± 0.07	0.99 ± 0.04
			HD95	56.00 ± 22.62	56.23 ± 24.09	57.18 ± 21.47	44.58 ± 26.80	54.91 ± 18.17
S_R – only	**0**	**60**	**DSC**	**0.98 ± 0.08**	**0.97 ± 0.17**	**0.93 ± 0.12**	**1.00 ± 0.00**	**0.97 ± 0.11**
			HD95	**23.04 ± 27.07**	**16.32 ± 24.15**	**10.92 ± 13.38**	**24.74 ± 37.29**	**36.40 ± 25.61**

Table 6
Communication costs for different compression pipelines. For each method, the minimum transmitted payload observed across training rounds is reported.

Pipeline	C	S_R (%)	Baseline bits	Compressed bits	Compression ratio	Payload reduction (%)
FedAvg	0	0	1.54×10^8	1.54×10^8	×1.00	0.00
$C-only$	3	0	1.54×10^8	4.09×10^7	×3.77	73.48
$C-only$	32	0	1.54×10^8	4.73×10^7	×3.26	69.31
$C-only$	64	0	1.54×10^8	4.91×10^7	×3.13	68.11
$S-only$	0	20	1.54×10^8	1.54×10^8	×1.00	0.00
$S-only$	0	30	1.54×10^8	1.53×10^8	×1.01	0.93
$S-only$	0	60	1.54×10^8	1.24×10^8	×1.24	19.48
FedWSOComp	32	60	1.54×10^8	4.52×10^7	×3.40	70.65

The communication results reported in Table 6 further clarified that the full FedWSOcomp pipeline represents a more practical operating point than clustering or sparsification alone. Although the $C-only$ configuration achieved substantial compression (up to 3.77×), the most aggressive configuration ($C = 3$) was associated with the weakest segmentation performance –presented in Tables 4 and 5– and the largest boundary errors.

Furthermore, the S_R-only configurations have not yielded consistent communication savings across retention levels while simultaneously increasing the boundary errors represented by the HD95.

The findings of this study, when considered in conjunction with the ablation and communication analyses, indicate that the primary advantage of FedWSOcomp arises from the synergistic combination of sufficiently fine-grained quantization with high retention, rather than from either compression mechanism in isolation. From a pragmatic standpoint, the configuration $C = 32$ with $S_R = 60\%$ was identified as the most balanced operating regime under IID conditions with constrained sample availability, while $C = 64$ with $S_R = 60\%$ was identified as the most effective configuration under non-IID settings. It was demonstrated that both configurations consistently achieved high DSC while maintaining low HD95, and simultaneously delivered high communication reductions without the instability associated with coarse quantization or low-retention sparsification.

Overall, these configurations delineate a stable and deployable trade-off between communication efficiency and segmentation reliability, thus supporting the assertion that FedWSOcomp may be considered suitable for clinical FL applications with constrained computational and network infrastructures, without compromising the robustness of tumour delineation.

6. Conclusion

This work analysed FedWSOComp, an FL strategy for communication-efficient 3D medical image segmentation, by systematically varying its key components, parameters and analysing their effects on model performance and compression. The study showed how different sparsification levels and quantization granularities influenced the trade-off between segmentation accuracy and communication efficiency, providing new insights into the practical applicability of the method. The results confirmed that moderate-to-fine quantization combined with higher retention ratios preserved segmentation quality while still achieving substantial communication savings. Moreover, the analysis of client-specific outcomes highlighted the importance of considering heterogeneity in federated scenarios, as improvements varied according to local data distributions. A limitation of this study concerned the construction of the IID scenario. By allocating 100 patients to each client, the total number of patients available in this case was lower than in the non-IID setting, which explained the performance degradation observed in some IID experiments. Overall, this study deepened the understanding of FedWSOComp and its operational boundaries, offering guidance for its deployment in bandwidth-constrained clinical networks. Future investigations may explore adaptive parameter selection, integration with privacy-preserving techniques, and large-scale validation across multiple institutions and imaging modalities.

CRediT authorship contribution statement

Asaf Raza: Writing – review & editing, Writing – original draft, Validation, Software, Methodology, Investigation, Formal analysis, Data curation. **Ciro Benito Raggio:** Writing – review & editing, Writing – original draft, Validation, Software, Methodology, Investigation, Formal analysis, Data curation. **Antonella Guzzo:** Writing – review & editing, Validation, Supervision, Methodology, Funding acquisition, Formal analysis, Conceptualization. **Maria Francesca Spadea:** Writing – review & editing, Validation, Supervision, Methodology, Formal analysis, Conceptualization. **Giancarlo Fortino:** Supervision, Project administration, Funding acquisition.

Funding

This work was partially supported by the following projects: (1) POS RADIOAMICA project funded by the Italian Ministry of Health (CUP: H53C22000650006); (2) PNRR project FAIR-Future AI Research (PE00000013), Spoke 9, under the NRRP MUR program funded by the NextGenerationEU.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability statement

The data used in this study were collected from the public Federated Tumor Segmentation (FeTS) Challenge 2024 dataset and are available at: https://doi.org/10.5281/zenodo.10990500.

References

[1] T.A. Soomro, L. Zheng, A.J. Afifi, A. Ali, S. Soomro, M. Yin, J. Gao, Image segmentation for mr brain tumor detection using machine learning: a review, IEEE Rev. Biomed. Eng. 16 (2022) 70–90, https://doi.org/10.1109/RBME.2022.3185292

[2] S. Karim, G. Tong, Y. Yu, A.A. Laghari, A.A. Khan, M. Ibrar, F. Mehmood, Developments in brain tumor segmentation using MRI: deep learning insights and future perspectives, IEEE Access 12 (2024) 26875–26896, https://doi.org/10.1109/ACCESS.2024.3365048

[3] S. Alphonse, F. Mathew, K. Dhanush, V. Dinesh, Federated learning with integrated attention multiscale model for brain tumor segmentation, Sci. Rep. 15 (2025) 11889, https://doi.org/10.1038/s41598-025-96416-6

[4] F.J. Dorfner, J.B. Patel, J. Kalpathy-Cramer, E.R. Gerstner, C.P. Bridge, A review of deep learning for brain tumor analysis in MRI, NPJ Precis. Oncol. 9 (2025) 2, https://doi.org/10.1038/s41698-024-00789-2

[5] O. Ronneberger, P. Fischer, T. Brox, U-Net: convolutional networks for biomedical image segmentation, in: Medical Image Computing and Computer-Assisted Intervention (MICCAI) of LNCS, vol. 9351, Springer, 2015, pp. 234–241, https://doi.org/10.1007/978-3-319-24574-4_28

[6] W. Zhang, W. Jin, S. Rho, F. Jiang, C.-F. Yang, A federated learning framework for brain tumor segmentation without sharing patient data, Int. J. Imaging Syst. Technol. 34 (2024) e23147, https://doi.org/10.1002/ima.23147

[7] K.S. Adewole, E. Alozie, H. Olagunju, N. Faruk, R.Y. Aliyu, A.L. Imoize, A. Abdulkarim, Y.O. Imam-Fulani, S. Garba, B.A. Baba, et al., A systematic review and meta-data analysis of clinical data repositories in Africa and beyond: recent development, challenges, and future directions, Discover Data 2 (2024) 8, https://doi.org/10.1007/s44248-024-00012-4

[8] H.B. McMahan, E. Moore, D. Ramage, S. Hampson, B.A. Y Arcas, Communication-efficient learning of deep networks from decentralized data, arXiv preprint arXiv:1602.05629, 2023.

[9] C. Shiranthika, P. Saeedi, I.V. Bajić, Decentralized learning in healthcare: a review of emerging techniques, IEEE Access 11 (2023) 54188–54209, https://doi.org/10.1109/ACCESS.2023.3281832

[10] C.B. Raggio, M.K. Zabaleta, N. Skupien, O. Blanck, F. Cicone, G.L. Cascini, P. Zaffino, L. Migliorelli, M.F. Spadea, FedSynthCT-Brain: a federated learning framework for multi-institutional brain MRI-to-CT synthesis, Comput. Biol. Med. 192 (2025) 110160, https://doi.org/10.1016/j.compbiomed.2025.110160

[11] C.B. Raggio, P. Zaffino, M.F. Spadea, A Privacy-Preserving federated learning framework for generalizable CBCT to synthetic CT translation in head and neck, arXiv preprint arXiv:2506.08654 2025

[12] S.S. Sandhu, H.T. Gorji, P. Tavakolian, K. Tavakolian, A. Akhbardeh, Medical imaging applications of federated learning, Diagnostics 13 (2023) 3140, https://doi.org/10.3390/diagnostics13193140

[13] C. Ren, H. Yu, H. Peng, X. Tang, B. Zhao, L. Yi, A.Z. Tan, Y. Gao, A. Li, X. Li, Z. Li, Q. Yang, Advances and open challenges in federated foundation models, IEEE Commun. Surv. Tutorials 28 (2026) 2087–2126, https://doi.org/10.1109/COMST.2025.3552524

[14] S. Gallus, A. Mercer, P. Singh, D. Cho, Optimization, communication, and personalization in federated learning for massive networks, Preprints (2025), https://doi.org/10.20944/preprints202507.1037.v1

[15] A. Yang, J. Liu, B. Yang, Z. Mo, K. Li, Balancing communication overhead and accuracy in compression integration: a survey, J. Supercomput. 81 (2025) 964, https://doi.org/10.1007/s11227-025-07451-z

[16] A.N. Sahu, A. Dutta, A.M. Abdelmoniem, T. Banerjee, M. Canini, P. Kalnis, Rethinking gradient sparsification as total error minimization, in: NeurIPS 2021 - Advances in Neural Information Processing Systems, 2021, https://arxiv.org/abs/2108.00951.

[17] D. Alistarh, D. Grubic, J.Z. Li, R. Tomioka, M. Vojnovic, Qsgd: communication-efficient SGD via gradient quantization and encoding, in: Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS'17, Curran Associates Inc., Red Hook, NY, USA, 2017, pp. 1707–1718.

[18] W. Wen, C. Xu, F. Yan, C. Wu, Y. Wang, Y. Chen, H. Li, TernGrad: ternary gradients to reduce communication in distributed deep learning, in: Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS'17, Curran Associates Inc., Red Hook, NY, USA, 2017, pp. 1508–1518.

[19] L. Wang, W. Wu, J. Zhang, H. Liu, G. Bosilca, M. Herlihy, R. Fonseca, FFT-based gradient sparsification for the distributed training of deep neural networks, in: Proceedings of the 29th International Symposium on High-Performance Parallel and Distributed Computing, HPDC '20, Association for Computing Machinery, New York, NY, USA, 2020, pp. 113–124, https://doi.org/10.1145/3369583.3392681

[20] A.N. Onaizah, Y. Xia, A.J. Obaid, K. Hussain, Deep learning based brain tumour architecture for weight sharing optimization in federated learning, Expert Syst. 42 (2025) e13643, https://doi.org/10.1111/exsy.13643

[21] P.S. Tejashwini, J. Thriveni, K.R. Venugopal, A novel SLCA-UNet architecture for automatic MRI brain tumor segmentation, Biomed. Signal Process. Control. 100 (2025) 107047, https://doi.org/10.1016/j.bspc.2024.107047

[22] K.G. Revathi, M.C.P. Shirley, S. Sreethar, et al., Brain tumor segmentation using optimized depth wise separable convolutional neural network with dense U-Net, Knowl.-Based Syst. (2025) 113678, https://doi.org/10.1016/j.knosys.2025.113678

[23] A. Shaikh, S. Amin, M.A. Zeb, A. Sulaiman, M.S. Al Reshan, H. Alshahrani, Enhanced brain tumor detection and segmentation using densely connected convolutional networks with stacking ensemble learning, Comput. Biol. Med. 186 (2025) 109703, https://doi.org/10.1016/j.compbiomed.2025.109703

[24] M. Obayya, A. Alshuhail, K. Mahmood, M.H. Alanazi, M. Alqahtani, N.O. Aljehane, H. Almansour, M.A. Al-Hagery, A novel u-net model for brain tumor segmentation from MRI images, Alex. Eng. J. 126 (2025) 220–230, https://doi.org/10.1016/j.aej.2025.04.051

[25] X. Guo, Z. Wang, P. Wu, Y. Li, F.E. Alsaadi, N. Zeng, Elts-net: an enhanced liver tumor segmentation network with augmented receptive field and global contextual information, Comput. Biol. Med. 169 (2024) 107879, https://doi.org/10.1016/j.compbiomed.2023.107879

[26] N. Zeng, H. Li, Z. Wang, W. Liu, S. Liu, F.E. Alsaadi, X. Liu, Deep-reinforcement-learning-based images segmentation for quantitative analysis of gold immunochromatographic strip, Neurocomputing 425 (2021) 173–180, https://doi.org/10.1016/j.neucom.2020.04.001, https://www.ciencedirect.com/science/article/pii/S0925231220305385.

[27] M.E. Yahiaoui, M. Derdour, R. Abdulghafor, S. Turaev, M. Gasmi, A. Bennour, A. Aborujilah, M. Sarem al, Federated learning with privacy preserving for multi-institutional three-dimensional brain tumor segmentation, Diagnostics 14 (2024) 2891, https://doi.org/10.3390/diagnostics14242891

[28] E. Albalawi, T.R. Mahesh, A. Thakur, V.V. Kumar, M. Gupta, S.B. Khan, A. Almusharraf, Integrated approach of federated learning with transfer learning for classification and diagnosis of brain tumor, BMC Med. Imaging 24 (2024) 110, https://doi.org/10.1186/s12880-024-01261-0

[29] J. Dai, H. Wu, H. Liu, L. Yu, X. Hu, X. Liu, D. Geng, FedATA: adaptive attention aggregation for federated self-supervised medical image segmentation, Neurocomputing 613 (2025) 128691, https://doi.org/10.1016/j.neucom.2024.128691

[30] R. Kumar, C.M. Bernard, A. Ullah, R.U. Khan, J. Kumar, D.K.B. Kulevome, R. Yunbo, S. Zeng, Privacy-preserving blockchain-based federated learning for brain tumor segmentation, Comput. Biol. Med. 177 (2024) 108646, https://doi.org/10.1016/j.compbiomed.2024.108646

[31] G. Appasami, N. Savarimuthu, Federated learning for secure medical MRI brain tumor image classification, Eur. Phys. J. Spec. Top. (2025) 1–15, https://doi.org/10.1140/epjs/s11734-025-01516-z

[32] J.B. Awotunde, C.O. Abikoye, B. Brahma, E.T. Oladipupo, A. Bandyopadhyay, Federated learning augmented with convolutional neural networks for brain cancer classification, Proc. Comput. Sci. 258 (2025) 2617–2626, https://doi.org/10.1016/j.procs.2025.04.523

[33] Z. Liu, F. Wu, Y. Wang, M. Yang, X. Pan, FedCL: federated contrastive learning for multi-center medical image classification, Pattern Recognit. 143 (2023) 109739, https://doi.org/10.1016/j.patcog.2023.109739

[34] A. Kazlouski, I.M. Perez, F. Noor, M. Högerman, O. Ettala, T. Pahikkala, A. Airola, Towards practical federated learning and evaluation for medical prediction models, Int. J. Med. Inform. (2025) 106046, https://doi.org/10.1016/j.ijmedinf.2025.106046

[35] P.N. Srinivasu, G.J. Lakshmi, S.C. Narahari, J. Shafi, J. Choi, M.F. Ijaz, Enhancing medical image classification via federated learning and pre-trained model, Egypt. Inform. J. 27 (2024) 100530, https://doi.org/10.1016/j.eij.2024.100530

[36] Y.N. Tan, V.P. Tinh, P.D. Lam, N.H. Nam, T.A. Khoa, A transfer learning approach to breast cancer classification in a federated learning framework, IEEE Access 11 (2023) 27462–27476, https://doi.org/10.1109/ACCESS.2023.3257562

[37] Y. Xu, J. Cheng, H. Xu, C. Guo, Y. Liao, Z. Yao, Towards layer-wise quantization for heterogeneous federated clients, Comput. Netw. 264 (2025) 111223, https://doi.org/10.1016/j.comnet.2025.111223

[38] M. Cao, H. Wang, Y. Yuan, J. Lu, X. Cai, D. Yu, M. Zhao, FedMQ +: towards efficient heterogeneous federated learning with multi-grained quantization, J. Syst. Archit. (2025) 103460, https://doi.org/10.1016/j.sysarc.2025.103460

[39] H. Hu, W. Du, Y. Li, Y. Wang, pFL-SBPM: a communication-efficient personalized federated learning framework for resource-limited edge clients, Future Gener. Comput. Syst. 171 (2025) 107849, https://doi.org/10.1016/j.future.2025.107849

[40] H. Zhang, T. Wu, Z. Ma, F. Li, J. Liu, Dynamic layer-wise sparsification for distributed deep learning, Futur. Gener. Comput. Syst. 147 (2023) 1–15, https://doi.org/10.1016/j.future.2023.04.022

[41] S. Bakas, A. Linardos, S. Pati, U. Baid, M. Sheller, B. Edwards, A. Karargyris, P. Mattson, B. Menze, M. Bilello, S. Mohan, J.B. Freymann, J.S. Kirby, C. Davatzikos, H. Fathallah-Shaykh, R. Wiest, A. Jakab, R.R. Colen, A. Kotrotsou, D. Marcus, M. Milchenko, A. Nazeri, M.-A. Weber, A. Mahajan, The federated tumor segmentation (FeTS) challenge 2024, Nat. Commun. (2024), https://doi.org/10.5281/ZENODO.10990500

[42] S. Bakas, et al., The multimodal brain tumor image segmentation benchmark (BRATS), IEEE Trans. Med. Imaging 41 (2022) 2773–2781, https://doi.org/10.1109/TMI.2022.3194052

[43] M.J. Cardoso, W. Li, R. Brown, N. Ma, E. Kerfoot, Y. Wang, B. Murrey, A. Myronenko, C. Zhao, D. Yang, V. Nath, Y. He, Z. Xu, A. Hatamizadeh, A.

Myronenko, W. Zhu, Y. Liu, M. Zheng, Y. Tang, I. Yang, M. Zephyr, B. Hashemian, S. Alle, M.Z. Darestani, C. Budd, M. Modat, T. Vercauteren, G. Wang, Y. Li, Y. Hu, Y. Fu, B. Gorman, H. Johnson, B. Genereaux, B.S. Erdal, V. Gupta, A. Diaz-Pinto, A. Dourson, L. Maier-Hein, P.F. Jaeger, M. Baumgartner, J. Kalpathy-Cramer, M. Flores, J. Kirby, L.A.D. Cooper, H.R. Roth, D. Xu, D. Bericat, R. Floca, S.K. Zhou, H. Shuaib, K. Farahani, K.H. Maier-Hein, S. Aylward, P. Dogra, S. Ourselin, A. Feng, MONAI: an open-source framework for deep learning in healthcare, arXiv preprint arXiv:2211.02701, 2022.

[44] P. Agrawal, N. Katal, N. Hooda, Segmentation and classification of brain tumor using 3d-unet deep neural networks, Int. J. Cogn. Comput. Eng. 3 (2022) 199–210, https://doi.org/10.1016/j.ijcce.2022.11.001, https://www.sciencedirect.com/science/article/pii/S2666307422000213.

[45] A.F. Aji, K. Heafield, Sparse communication for distributed gradient descent, in: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, 2017, https://doi.org/10.18653/v1/d17-1045

[46] D.J. Beutel, T. Topal, A. Mathur, X. Qiu, J. Fernandez-Marques, Y. Gao, L. Sani, K.H. Li, T. Parcollet, P.P.B. de Gusmão, N.D. Lane, Flower: A friendly federated learning research framework, arXiv preprint arXiv:2007.14390, 2022.

[47] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Köpf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, S. Chintala, PyTorch: an imperative style, High-Performance deep learning library, arXiv preprint arXiv:1912.01703, 2019.

[48] A. Celaya, B. Riviere, D. Fuentes, A generalized surface loss for reducing the hausdorff distance in medical imaging segmentation, arXiv preprint arXiv:2302.03868, 2024

[49] J. Li, Y. Zhang, Y. Li, X. Gong, W. Wang, Fedsparse: a communication-efficient federated learning framework based on sparse updates, Electronics 13 (2024) 5042, https://doi.org/10.3390/electronics13245042

[50] J. Jia, J. Liu, C. Zhou, H. Tian, M. Dong, D. Dou, Efficient asynchronous federated learning with sparsification and quantization, Concurr. Comput. Pract. Exp. 36 (2024) e8002, https://doi.org/10.1002/cpe.8002

Author biography

Ciro Benito Raggio, M.Sc., is a Research Associate at the Karlsruhe Institute of Technology, Germany, where he is pursuing a PhD in Decentralized AI for Healthcare and Medical Imaging. He received his B.Sc. in Computer Science and Biomedical Engineering and his M.Sc. in Biomedical Engineering from the Magna Graecia University of Catanzaro, Italy. His research interests include federated learning and privacy-preserving AI for decentralized medical image-to-image translation, with applications in synthetic CT generation and computer-aided diagnosis. He has contributed to open-source tools, peer-reviewed publications, and international AI collaborations.

Asaf Raza received his PhD in Information and Communication Technologies from the University of Calabria (UniCal), Italy, where his research focused on Federated Learning and Radiomics for privacy-preserving medical image analysis, particularly in brain tumor imaging. He previously earned his Master's degree in Software Engineering from the University of Engineering and Technology (UET) Taxila, Pakistan, with a research focus on brain tumor classification using deep learning. His research interests include federated learning, radiomics, and AI-driven medical imaging.

Antonella Guzzo is an Associate Professor of Computer Engineering at the University of Calabria, Italy. She holds a Ph.D. in Computer Engineering (2005) and previously worked as a research fellow at ICAR-CNR. Her research focuses on process mining, data mining, AI, and deep learning. She has published over 75 papers, serves as a reviewer for numerous international journals, and is active in program committees for major conferences. Since 2005, she has contributed to several national and international research projects. She is a cofounder of the IEEE Task Force on Process Mining and a member of the IEEE Working Group for the XES standard.

Maria Francesca Spadea is Chair of Modelling and Simulation for Medical Engineering at Karlsruhe Institute of Technology (Germany), where she co-leads the Institute of Biomedical Engineering. She earned her MSc in Mechanical Engineering from the University of Calabria (2002) and her PhD in Bioengineering from Politecnico di Milano (2006). Between 2006 and 2015, she was a Visiting Professor at the Massachusetts Institute of Technology and Harvard Medical School, where she was also a Fulbright Research Scholar (2010–2011), focusing on metal artefact reduction in CT imaging. In the past 15 years, she has established extensive national and international collaborations across academia, clinical institutions, and industry. Her research focuses on developing and validating computational methods to support diagnosis and therapy, including medical image processing and analysis, surgical navigation and advanced image guided radiotherapy.

Giancarlo Fortino is a Full Professor of Computer Engineering at the University of Calabria, Italy, and a distinguished professor at multiple Chinese universities. He is a senior research fellow at ICAR-CNR, a CAS PIFI visiting scientist, and an IEEE Sensors Council Distinguished Lecturer. His research focuses on wearable systems, e-health, IoT, agent-based computing, and AI-driven human–machine systems. Prof. Fortino has chaired 130+ conferences/workshops, served on 700+ program committees, and guest-edited over 80 special issues. He has also organized numerous high-impact events, including recent workshops on GenAI for IoT (WF-IoT 2024), vehicle formation control (IEEE CASE 2024), and distributed collective intelligence (ICDCS 2022).