Natural Resources for Human Health

eISSN: 2583-1194

Original Research

Received 15 March 2026
Revised 20 April 2026
Accepted 10 May 2026

Natural Resources for Human Health 2026, Vol. 6 (1s): 12–38

KEYWORDS:
Federated learning;
medical image classification;
non-IID data;
attention mechanisms;
metaheuristic feature selection;
explainable artificial intelligence;
blockchain audit trail;

Fed-TransResCaps-HGBOA-GWO-PSO-Net: A Secure and Explainable Federated Hybrid Deep Learning Framework for CT–MRI Medical Image Classification

Om Prakash Karada[1], Prashant Panse[2]

[1]PhD Scholar, Medicaps University, Indore, Madhya Pradesh, India
[2]Professor, Medicaps University, Indore, Madhya Pradesh, India
[1]opkarada@gmail.com, [2]prashantpanse@gmail.com

Corresponding author: Om Prakash Karada, Medicaps University, Indore, India

Abstract: Hospitals rarely share raw imaging data. Privacy statutes, institutional review requirements, and plain commercial caution keep scans behind institutional firewalls, which leaves most large-scale medical imaging models trained on whatever single-site corpus their authors could assemble. Federated learning offers a way out, but it inherits two problems that centralised training never had to face: client data are not identically distributed, and the model updates that cross the network are themselves a disclosure surface. This paper presents Fed-TransResCaps-HGBOA-GWO-PSO-Net, a federated framework for CT–MRI modality classification that addresses both. Three complementary encoders a frozen ResNet50, a frozen EfficientNetB0, and a lightweight convolutional-transformer branch are refined by Convolutional Block Attention Modules and concatenated into a 3392-dimensional descriptor. A hybrid Honey-Badger/Grey-Wolf/Particle-Swarm search then prunes this descriptor to a 70% subset, so that only 2374 coordinates reach the classification head and the per-round payload shrinks accordingly. Clients are grouped by label distribution using k-means before aggregation, and updates are combined with FedNova first within and then across clusters, which keeps a pathologically skewed client from dominating the global model. Every accepted update is committed to a SHA-256 hash chain, giving an append-only audit trail of who contributed what and when. Grad-CAM overlays are produced at inference so that a clinician can see which image regions drove a decision. On a 4974-image CT–MRI benchmark partitioned across three deliberately non-IID clients, the framework converged from 56.38% to 98.66% accuracy in twelve communication rounds, with an AUC of 0.9979, specificity of 0.9973, and Matthews correlation of 0.9733. The dominant residual error is a one-sided 2.42% false-negative rate on MRI, which we trace to the label-skewed client partition rather than to the encoder.

1 INTRODUCTION

Computed tomography and magnetic resonance imaging are the two workhorses of neurological diagnosis, and the choice between them is rarely arbitrary. CT resolves bone and acute haemorrhage quickly and cheaply; MRI resolves soft tissue, oedema and tumour margins with a contrast range CT cannot approach. A modern radiology archive therefore holds both, often for the same patient, and any automated system that reads from such an archive has to know which modality it is looking at before it can do anything else. Modality-aware routing, cross-modal registration, quality control on ingest, and the curation of training corpora for downstream diagnostic models all depend on that determination being correct, and all of them break silently when it is not.

Deep learning has largely solved this class of problem in the laboratory. Convolutional networks, transfer-learned backbones, vision transformers and hybrid attention architectures all report high accuracy on medical image classification benchmarks [1], [3], [4], [5]. What has not been solved is getting such a system trained on data that reflects more than one institution. Imaging archives are held behind hospital firewalls for good reasons patient privacy legislation, institutional review requirements, and the plain commercial reality that scans are an asset. The consequence is that most published medical imaging models are trained on whatever single-site corpus their authors could obtain, and their reported performance is an estimate of how well they work at that site rather than anywhere else.

Federated learning was proposed to break this deadlock by moving the model to the data instead of the reverse [9], [13], [14]. Clients train locally and transmit only parameters; a coordinating server aggregates them into a shared model. The idea is now well established in medical imaging, with demonstrations spanning classification [10], [12], [15], segmentation [17], [19], [20] and reconstruction [6]. But federation trades one hard problem for two others, and both bear directly on the design presented here.

The first is statistical heterogeneity. Client datasets are not independent and identically distributed: hospitals differ in scanner manufacturer, field strength, reconstruction kernel, acquisition protocol, referral pattern and case mix. When local label distributions diverge, so do local optima, and a naive parameter average lands at a point that is optimal for no participant. Drift-corrected aggregation schemes address this with control variates [23] or by normalising each client's accumulated update [27], but the pathology is severe enough under strong label skew that averaging alone is insufficient a point our own round-one measurement makes vividly, and which we return to in Section 4.

The second is that the transmitted updates are themselves a disclosure surface, and a weaker one than is often assumed. Withholding raw pixels is not the same as withholding information about them. Secure aggregation and differential privacy constrain what an update reveals [7], [31], but neither produces a durable record of which participant contributed which parameters at which round and that record, not confidentiality alone, is what an institutional audit actually requires.

Two further limitations run through the literature. Feature selection is almost universally absent from federated pipelines, even though the transmitted payload scales with the representation reaching the classification head, which makes selection considerably more valuable in a federated setting than in a centralised one. And interpretability, where present at all, is bolted on as a visualisation step [8] rather than treated as one half of a single accountability requirement whose other half is provenance.

This paper addresses all four concerns in one framework. Fed-TransResCaps-HGBOA-GWO-PSO-Net encodes each scan through three complementary branches, refines two of them with convolutional block attention, prunes the fused descriptor with a hybrid metaheuristic search, aggregates client updates hierarchically after grouping clients by label distribution, commits every accepted update to a SHA-256 hash chain, and produces Grad-CAM overlays at inference. It is evaluated on a deliberately adversarial three-client federation in which two of the three clients hold a single modality each.

1.1 Contributions

1. **A three-branch encoder with attention refinement.** Frozen ResNet50 and EfficientNetB0 backbones supply complementary convolutional representations and a lightweight convolutional-transformer branch supplies global context; CBAM refines both convolutional streams along the channel and spatial axes. Freezing the backbones fixes an identical prior at every client, removing one source of inter-client drift before aggregation begins, and reduces the trainable parameter count to 2.38 M of 30.02 M.
2. **Metaheuristic feature selection inside the federated loop.** A composed Honey-Badger, Grey-Wolf and Particle-Swarm search selects a fixed binary mask retaining 70% of the 3392-dimensional fused descriptor. Because the mask is computed once and shared, head weights remain comparable across sites, and the retained width determines the per-round payload for every client in every round.
3. **Distribution-space clustering with two-level FedNova aggregation.** Clients are grouped by label distribution using k-means before training, then aggregated within and across clusters. Grouping compatible clients first means only partially reconciled cluster representatives meet at the top level, which is what allows the framework to recover from a first-round collapse to 56.38% within two rounds.

4. **An append-only integrity chain with an explicit threat model.** Every accepted update is committed to a SHA-256 chain recording round, client, cluster, shard size, local accuracy and a parameter digest. We state precisely what this provides tamper-evidence over training history and what it does not: it is an audit mechanism, not a defence against poisoning, and it offers no confidentiality guarantee for the parameters themselves.
5. **Evaluation that reports where the errors are, not only how many.** Beyond aggregate accuracy, precision, recall, F1-score, AUC and MCC, we decompose performance per class, per client and per round, and trace the residual error to a specific property of the partition rather than to the encoder. We also report an analytical cost decomposition showing that the attention branch consumes roughly 96% of the forward pass while contributing under 2% of the fused descriptor a finding that argues against the branch as currently configured.

The remainder of the paper is organised as follows. Section 2 reviews the three literatures the work draws on and states the gap it occupies. Section 3 develops the framework, its mathematical formulation and its five constituent algorithms. Section 4 reports the experimental results, the systems-level measurements and the limitations that bound their interpretation.

2 LITERATURE REVIEW

Work relevant to this paper falls into three streams that have, until recently, developed largely in parallel: architectural advances in medical image classification, systems work on federated optimisation under heterogeneity, and the interpretability and integrity mechanisms that make a distributed clinical model auditable. We take each in turn and then state what remains unresolved at their intersection.

2.1 Architectural Progress in Medical Image Analysis

Convolutional networks established the performance baseline for radiological classification and remain the default choice for single-site work [1], [3]. The most consistent gains since then have come not from depth but from selective emphasis: attention-embedded CNNs with residual shortcuts improve multiclass tumour discrimination by weighting informative channels and suppressing background response [2], and broader surveys of AI-driven diagnosis reach the same conclusion across modalities and tasks [4].

The shift toward transformers changed which errors these models make rather than simply reducing their count. Self-attention captures long-range dependencies that a stack of 3×3 kernels reaches only through many layers, and hybrid CNN–transformer designs applied to 3D MRI report better boundary delineation as a result [5], [25]. That gain is not free. Attention is quadratic in token count, and a naive transformer over a 128×128 feature map is considerably more expensive than the convolutional trunk it augments a real constraint when the model must eventually run on hospital-grade hardware rather than a dedicated cluster.

Two mitigations recur in the literature. The first is to restrict attention to a compact learned embedding rather than the full feature map, which is the approach we adopt in Section 3. The second is to compose several pretrained backbones and fuse their outputs, accepting a larger parameter count in exchange for shorter training and better generalisation from limited annotation [11], [30]. Multi-model fusion with attention-enhanced CNNs has been reported to outperform any single constituent in decentralised settings [11], though the fusion literature is uneven about whether the improvement survives a properly matched parameter budget. Figure 1 places these architectural families against the pipeline stages they occupy.

Natural Resources for Human Health
eISSN: 2583-1194

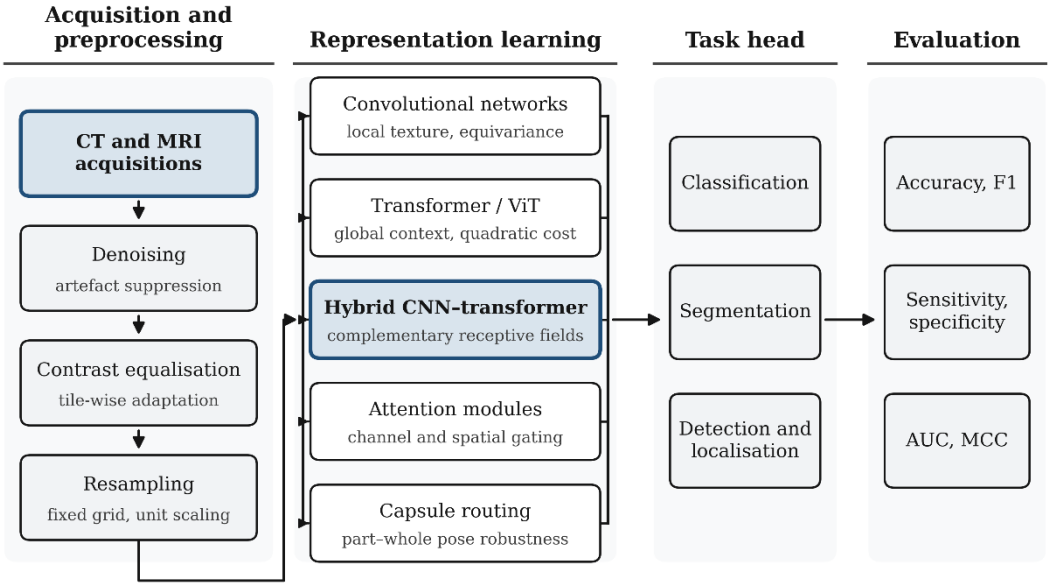


Figure 1: Architectural families for medical image analysis and the pipeline stages they occupy. Capacity increases from left to right within the representation-learning column, and so does data demand and inference cost; the hybrid CNN-transformer configuration adopted in Section 3 sits in the middle of that trade-off.

2.2 Federated Optimisation under Client Heterogeneity

Federated learning exists precisely so that institutions can train jointly without moving records, and scalable formulations of it have been demonstrated on brain imaging tasks with explicit privacy provisions [13]. Its clinical appeal is obvious enough that adoption in neuroimaging followed quickly [9], [14], [15]. Early medical deployments largely reused FedAvg with transfer-learned backbones [10], [12], [16], [28], [29], which works acceptably when client distributions are close to identical and degrades sharply when they are not.

The degradation has a well-understood cause. When clients hold different label distributions, their local optima diverge, and a naive parameter average lands somewhere between them that is optimal for no one. SCAFFOLD addresses this with control variates [23]; FedProx constrains local drift with a proximal term; FedNova normalises each client's accumulated update by its number of local steps so that clients performing more work do not implicitly acquire more influence. Comparative work on non-IID partitions in medical imaging generally favours the drift-corrected variants [27], and reviews of MRI-based federation identify aggregation strategy as the single most consequential design decision [31].

Clustered federation is a complementary idea that has received less attention in the clinical literature than it deserves. Rather than reconciling all clients into one global model, clients with similar distributions are grouped and aggregated within group before a second-level combination. The intuition is that averaging is well behaved among similar clients and poorly behaved among dissimilar ones, so the hierarchy should follow the similarity structure. Related decentralised efforts in segmentation [17], [20], stroke lesion modelling [24], and BraTS-derived federations [19] report the expected benefit, and edge-oriented deployments show that the reduced synchronisation also lowers latency [18], [26]. Capsule-based federated classifiers have been explored for the same reason: part–whole routing is more robust to the pose and intensity variation that distinguishes one site's scanner from another's [22]. Figure 2 summarises the canonical arrangement and the problems that remain open within it.

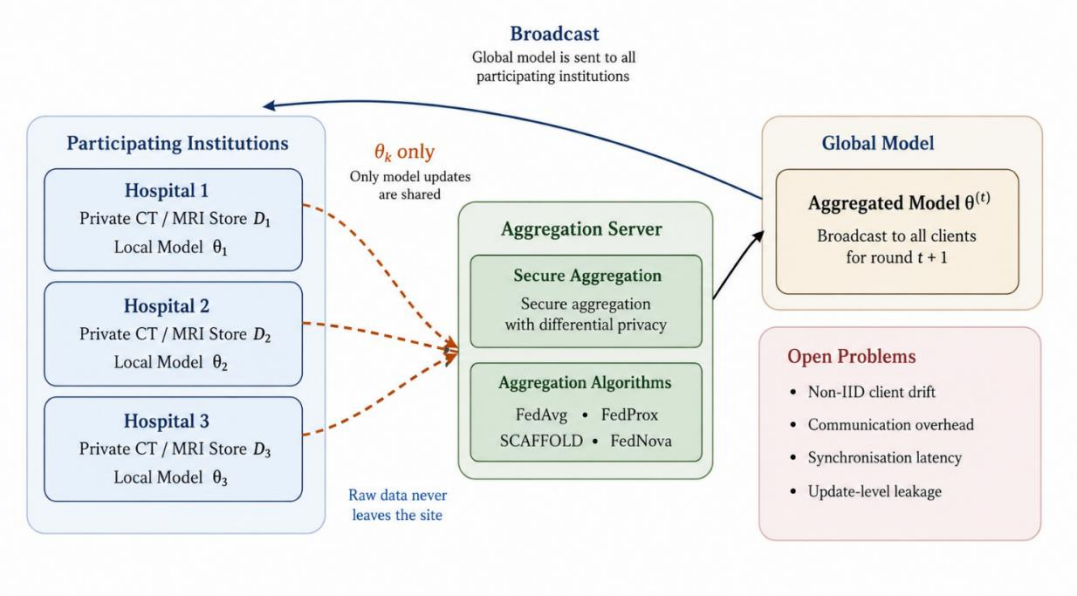


Figure 2: Canonical federated learning arrangement for multi-institutional medical imaging. Only parameters cross the network; the aggregation rule at the server is the design decision with the largest effect on behaviour under heterogeneity. The four items on the right are the failure modes this paper addresses.

2.3 Interpretability, Integrity, and What Remains Open

A federated clinical model raises two questions that a centralised one does not. A clinician asks why the model decided as it did; a compliance officer asks who contributed which update. Grad-CAM and SHAP have become the default answer to the first, and have been demonstrated inside federated pipelines for multi-institutional tumour diagnosis [7], [8]. Attention maps produced for accuracy reasons are often pressed into service as a coarse interpretability signal as well, and federated diagnostic frameworks for other neurological conditions report the same tension between predictive performance and auditability [32]. The second question is less well served. Secure aggregation and differential privacy protect update contents, but neither produces a durable record of provenance, which is what an audit actually requires. Cryptographic hash chaining is the natural fit and is beginning to appear in federated designs, though rarely with any analysis of what it does and does not guarantee. Figure 3 separates the two concerns.

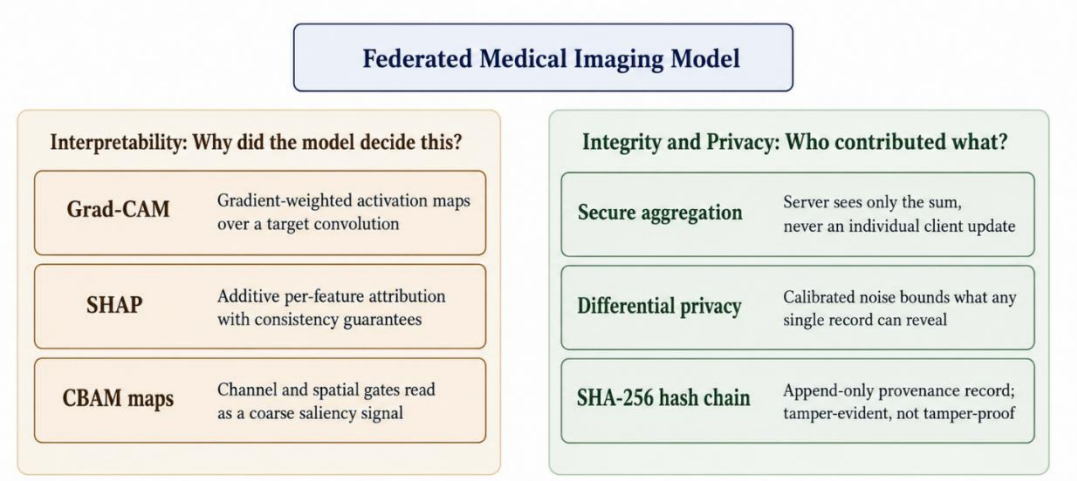


Figure 3: Interpretability and integrity are distinct requirements served by distinct mechanisms. The left column answers a clinician's question about a single prediction; the right column answers an auditor's question about the training history. A system that provides only one of the two is incomplete for clinical deployment.

Three gaps follow from this survey. First, the architectural literature and the federated-optimisation literature have advanced separately: hybrid attention encoders are usually evaluated centrally, and aggregation strategies are usually tested on small backbones. Second, feature selection is almost universally omitted from federated pipelines, even though the transmitted payload scales with the representation that reaches the head, making selection unusually valuable here compared with centralised training. Third, interpretability and provenance are treated as separate add-ons rather than as a single accountability requirement. Figure 4 maps these and related open problems onto the directions the field is pursuing, and Table 1 summarises where representative prior work sits against the three criteria above; the framework developed in Section 3 is designed to occupy the row that none of them fills.

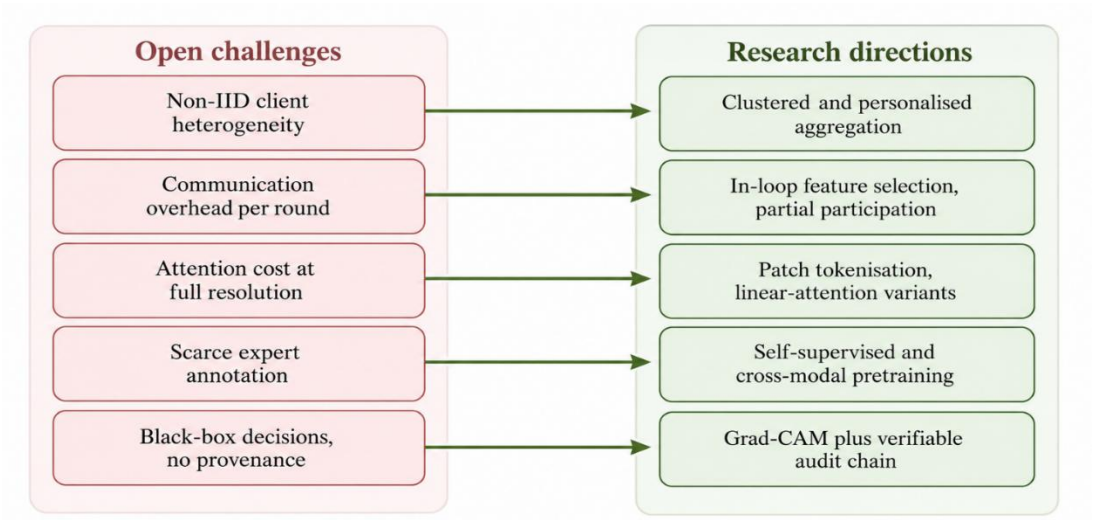


Figure 4: Open challenges in federated medical imaging and the corresponding research directions. Each row pairs a limitation identified in this review with the response adopted in Section 3.

Table 1: Positioning of representative prior work against the three criteria identified in Section 2. "Partial" indicates the capability is present but not evaluated as a contribution.

Study	Hybrid attention encoder	Heterogeneity-aware aggregation	Feature selection in the loop	Interpretability	Provenance record
Attention CNN [2]	Yes	Not federated	No	Partial	No
CNN–ViT hybrid [5], [25]	Yes	Not federated	No	No	No
FedAvg + transfer [10], [12]	No	No	No	No	No
DenseFedCNN [27]	Partial	Yes	No	No	No
SCAFFOLD federation [23]	No	Yes	No	No	No
Federated CapsNet [22]	Partial	Partial	No	No	No
Explainable federation [7], [8]	Partial	No	No	Yes	No

Study	Hybrid attention encoder	Heterogeneity-aware aggregation	Feature selection in the loop	Interpretability	Provenance record
Federated CBAM-VGG19 [32]	Yes	No	No	Partial	No
This work	**Yes**	**Yes (clustered FedNova)**	**Yes (HGBOA–GWO–PSO)**	**Yes (Grad-CAM)**	**Yes (SHA-256 chain)**

3 PROPOSED FRAMEWORK

3.1 Architectural Overview

The framework is organised as four stages that map onto the physical boundaries of a federated deployment: what happens inside a hospital, what leaves it, what the coordinating server does, and what is returned. The site-local half of that pipeline is shown first, in Figure 5, because it is the part a participating institution has to implement and audit for itself.

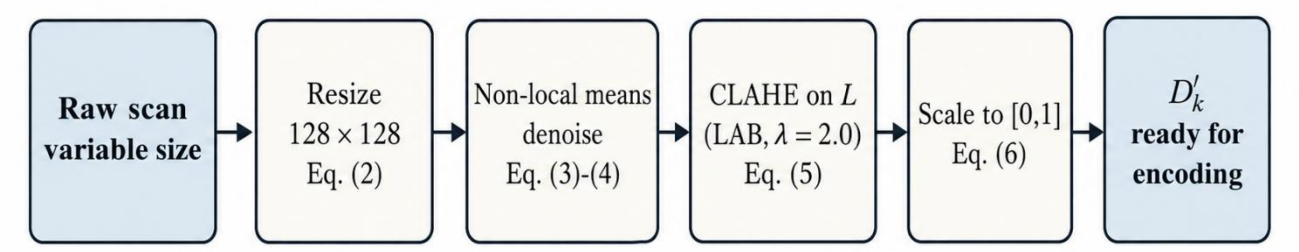


Figure 5: Site-local preprocessing chain. Every participating institution applies this sequence with identical hyperparameters before any tensor is formed; raw pixels are consumed here and never transmitted.

Figure 6 shows how the four stages fit together.

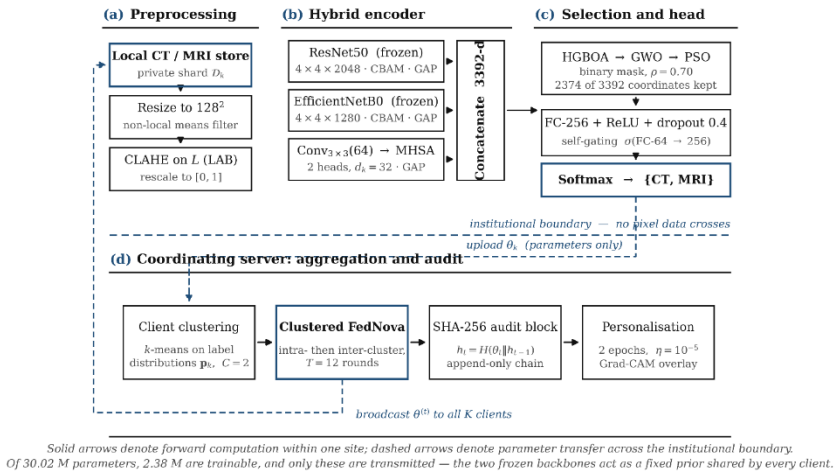


Figure 6: Fed-TransResCaps-HGBOA-GWO-PSO-Net. Stage (a) runs entirely inside the participating institution and never emits pixels. Stage (b) encodes each image through three complementary branches, two of which are refined by CBAM. Stage (c) prunes the fused descriptor with the hybrid metaheuristic and applies the local classification head. Stage (d) resides on the coordinating server, where updates are clustered, aggregated with FedNova at two levels, committed to a SHA-256 chain, and returned. Solid arrows denote computation; dashed arrows denote transfers across the network.

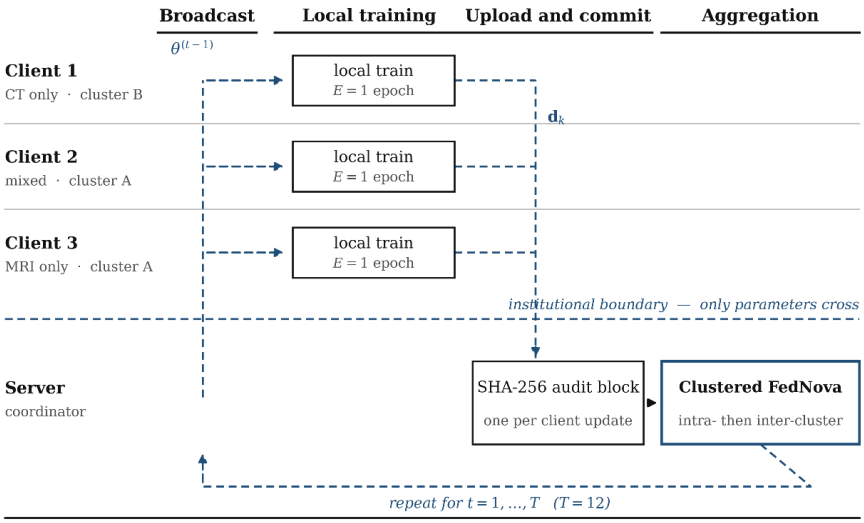
Stage (a) is preprocessing, and it is deliberately identical at every site. Each scan is resized to 128×128, denoised with non-local means, contrast-equalised with CLAHE applied to the luminance channel of the LAB representation, and scaled to the unit interval. Fixing this pipeline across clients matters more in a federated setting than a centralised one: a preprocessing discrepancy between sites is indistinguishable, from the server's point of view, from a genuine distribution shift, and the aggregation machinery will try to correct for it.

Stage (b) encodes the preprocessed image three ways. A frozen ResNet50 supplies deep residual features; a frozen EfficientNetB0 supplies a compound-scaled representation at roughly a fifth of the parameter cost; and a lightweight convolutional-transformer branch supplies global context that neither convolutional trunk captures well. Both convolutional branches pass through CBAM, which applies channel attention followed by spatial attention. Freezing the two ImageNet backbones is a deliberate choice rather than a computational shortcut. With 2.38M trainable parameters out of 30.02M total, the per-round communication payload is determined by the trainable fraction alone, and the frozen weights act as a fixed, identical prior at every client which removes one source of inter-client drift before aggregation ever begins.

Stage (c) is where the framework departs most sharply from standard federated pipelines. The three branch outputs concatenate into a 3392-dimensional vector, and a hybrid Honey-Badger/Grey-Wolf/Particle-Swarm search selects a binary mask retaining 70% of these coordinates. In centralised training, feature selection at this point buys a modest regularisation benefit. In a federated system it also reduces the head's input width, and therefore the size of every update that subsequently crosses the network, in every round, for every client.

Stage (d) is the server. Clients are first grouped by label distribution using k-means, then aggregated with FedNova within each cluster and again across clusters. Each accepted update is committed to a SHA-256 hash chain before the global model is broadcast. After the final round, each client fine-tunes the global model on its own data for two epochs, giving a personalised model alongside the shared one.

The temporal view of a single round is easier to read as a lane diagram than as a block diagram, and Figure 7 gives it. What crosses the network in either direction is a parameter tensor; the ordering constraint that matters is that the hash commit happens before the aggregation, so that the audit record reflects what each client actually submitted rather than what survived averaging.



Clients 2 and 3 share a distribution cluster and are reconciled first; client 1, holding a single modality, meets them only at the inter-cluster step. The audit block is written before aggregation, so the record reflects what each client submitted rather than what survived averaging.

Figure 7: Timeline of one global communication round. Dashed arrows are network transfers, solid boxes are local computation. Clients 2 and 3 fall in one distribution cluster and client 1 in the other, so intra-cluster aggregation resolves the two single-class extremes against the mixed shard before the inter-cluster step.

3.2 Preprocessing and Data Representation

Let the union of all participating institutions' data be

$$\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{N}, \qquad y_i \in \{0,1\}$$

where $\mathbf{x}_i$ is the i-th scan, y_i encodes the acquisition modality with 0 for CT and 1 for MRI, and N is the total sample count. This set is never materialised anywhere; it is a notational device, and the entire point of the framework is that no participant observes more than its own share of it.

Each image is resized to a common resolution:

$$\mathbf{x}_i^r = \mathcal{R}(\mathbf{x}_i,\, 128 \times 128)$$

Acquisition noise is suppressed by non-local means, which replaces each pixel with a similarity-weighted average over the search window Ω rather than a spatial neighbourhood:

$$\mathbf{x}_i^d(p) = \frac{1}{Z(p)} \sum_{q \in \Omega} \exp\left(-\frac{\|\mathcal{N}_p - \mathcal{N}_q\|_{2,a}^2}{h^2} \right) \mathbf{x}_i^r(q)$$

$$Z(p) = \sum_{q \in \Omega} \exp\left(-\frac{\|\mathcal{N}_p - \mathcal{N}_q\|_{2,a}^2}{h^2} \right)$$

Here $\mathcal{N}_p$ is the patch centred at p, $\|\cdot\|_{2,a}$ is a Gaussian-weighted Euclidean distance with standard deviation a, and h controls filter strength. Because the weighting is driven by patch similarity rather than proximity, edges and fine texture survive the operation — which is the property that matters when the discriminative signal is textural, as it is for modality identification.

Local contrast is then equalised. The image is converted to LAB space and CLAHE is applied to the luminance channel only, leaving chrominance untouched:

$$\mathbf{x}_i^c = \Psi^{-1}\left(\mathrm{CLAHE}_{\lambda}\left(L\left(\Psi(\mathbf{x}_i^d) \right) \right),\ a, b \right)$$

where Ψ denotes RGB-to-LAB conversion and λ is the clip limit. CLAHE partitions the image into 8×8 tiles and equalises within each, redistributing histogram mass above λ uniformly so that homogeneous tiles do not have their noise amplified. Intensities are finally normalised:

$$\mathbf{x}_i^n = \frac{\mathbf{x}_i^c}{255}, \qquad \mathbf{x}_i^n \in [0,1]^{128 \times 128 \times 3}$$

giving the preprocessed corpus

$$\mathcal{D}' = \{(\mathbf{x}_i^n, y_i)\}_{i=1}^{N}$$

3.3 Non-IID Client Construction and Clustering

The corpus is partitioned across K clients under a label-sorted scheme that produces a deliberately adversarial heterogeneity profile:

$$\mathcal{D}' = \bigcup_{k=1}^{K} \mathcal{D}_k, \qquad \mathcal{D}_j \cap \mathcal{D}_k = \varnothing \ \ \forall j \neq k, \qquad \sum_{k=1}^{K} n_k = N$$

with $n_k = |\mathcal{D}_k|$. Each client's empirical label distribution is

$$p_k(c) = \frac{1}{n_k} \sum_{(\mathbf{x},y) \in \mathcal{D}_k} \mathbb{1}[y = c], \qquad \mathbf{p}_k = [p_k(0),\, p_k(1)]^{\top}$$

Heterogeneity severity is quantified by the Jensen–Shannon divergence between each client and the global distribution $\bar{\mathbf{p}}$:

$$\delta_k = \mathrm{JSD}(\mathbf{p}_k \parallel \bar{\mathbf{p}}) = 1/2\,\mathrm{KL}(\mathbf{p}_k \parallel \mathbf{m}) + 1/2\,\mathrm{KL}(\bar{\mathbf{p}} \parallel \mathbf{m}), \qquad \mathbf{m} = 1/2\,(\mathbf{p}_k + \bar{\mathbf{p}})$$

Clients are then grouped into C clusters by minimising within-cluster dispersion over these distribution vectors:

$$\mathcal{C}^* = \underset{\mathcal{C}}{\operatorname{argmin}} \sum_{c=1}^{C} \sum_{k \in \mathcal{C}_c} \|\mathbf{p}_k - \boldsymbol{\mu}_c\|_2^2, \qquad \boldsymbol{\mu}_c = \frac{1}{|\mathcal{C}_c|} \sum_{k \in \mathcal{C}_c} \mathbf{p}_k$$

Clustering on $\mathbf{p}_k$ rather than on model parameters is a design decision worth stating explicitly. Parameter-space clustering requires a full round of training before the grouping can be computed and must be recomputed as models drift; distribution-space clustering is available before training starts, costs nothing, and for label skew, which is the dominant heterogeneity mode in multi-institutional imaging captures the structure that actually matters.

3.4 Hybrid Three-Branch Encoding

Residual branch. The ResNet50 trunk composes bottleneck units of the form

$$\mathbf{h}^{(l+1)} = \sigma_{\mathrm{ReLU}}\left(\mathcal{F}\left(\mathbf{h}^{(l)}, \left\{\mathbf{W}_j^{(l)}\right\}\right) + \mathcal{S}\left(\mathbf{h}^{(l)}\right)\right)$$

where $\mathcal{F}$ is the residual mapping and $\mathcal{S}$ is identity or a projection when dimensions change. The final stage yields $\mathbf{F}_R \in \mathbb{R}^{4\times4\times2048}$.

Compound-scaled branch. EfficientNetB0 balances depth d, width w, and resolution r under a single coefficient ϕ:

$$d = \alpha^{\phi}, \quad w = \beta^{\phi}, \quad r = \gamma^{\phi} \qquad \text{subject to} \quad \alpha \cdot \beta^2 \cdot \gamma^2 \approx 2, \quad \alpha, \beta, \gamma \geq 1$$

producing $\mathbf{F}_E \in \mathbb{R}^{4\times4\times1280}$ at roughly one-fifth the parameter cost of the residual trunk. The two branches are retained together because they fail differently: the residual trunk is the stronger texture encoder, while the compound-scaled trunk is more stable under the intensity shifts that separate scanners.

Transformer branch. Rather than attending over a full backbone feature map, the input is first projected to a compact token sequence:

$$\mathbf{T}^{(0)} = \mathrm{Reshape}\left(\sigma_{\mathrm{ReLU}}(\mathbf{W}_c * \mathbf{x}_i^n + \mathbf{b}_c)\right) \in \mathbb{R}^{L\times d_m}, \qquad L = 128^2,\ d_m = 64$$

Scaled dot-product attention is applied over these tokens:

$$\mathrm{Attn}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \mathrm{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^{\top}}{\sqrt{d_k}}\right)\mathbf{V}, \qquad \mathbf{Q} = \mathbf{T}^{(0)}\mathbf{W}^{Q},\ \mathbf{K} = \mathbf{T}^{(0)}\mathbf{W}^{K},\ \mathbf{V} = \mathbf{T}^{(0)}\mathbf{W}^{V}$$

with $H = 2$ heads of dimension $d_k = 32$ concatenated and projected:

$$\mathrm{MHSA}\left(\mathbf{T}^{(0)}\right) = [\mathrm{head}_1 \oplus \cdots \oplus \mathrm{head}_H]\mathbf{W}^{O}, \qquad \mathrm{head}_h = \mathrm{Attn}(\mathbf{Q}_h, \mathbf{K}_h, \mathbf{V}_h)$$

$$\mathbf{f}_T = \frac{1}{L}\sum_{\ell=1}^{L} \mathrm{MHSA}\left(\mathbf{T}^{(0)}\right)_{\ell,:} \in \mathbb{R}^{64}$$

3.5 Convolutional Block Attention

Both convolutional branches are refined by CBAM. Channel attention aggregates each map by average and max pooling, passes both through a shared bottleneck MLP with reduction ratio $r = 8$, and sums the responses:

$$\mathbf{M}_c(\mathbf{F}) = \sigma\left(\mathbf{W}_1\sigma_{\mathrm{ReLU}}\left(\mathbf{W}_0\,\mathrm{AvgPool}(\mathbf{F})\right) + \mathbf{W}_1\sigma_{\mathrm{ReLU}}\left(\mathbf{W}_0\,\mathrm{MaxPool}(\mathbf{F})\right)\right)$$

$$\mathbf{F}' = \mathbf{M}_c(\mathbf{F}) \otimes \mathbf{F}$$

where σ is the logistic function and $\otimes$ is channel-wise broadcast multiplication. Spatial attention then reduces along the channel axis and convolves the two resulting maps:

$$\mathbf{M}_s(\mathbf{F}') = \sigma\big(f^{7\times7}([\mathrm{AvgPool}_{ch}(\mathbf{F}')\,;\,\mathrm{MaxPool}_{ch}(\mathbf{F}')])\big)$$

$$\mathbf{F}'' = \mathbf{M}_s(\mathbf{F}') \otimes \mathbf{F}'$$

The order is not arbitrary: channel attention first establishes *what* is informative, and spatial attention then localises *where*, so the spatial map is computed over an already-reweighted volume.

3.6 Feature Fusion

Global average pooling reduces each refined convolutional volume to a vector, which is concatenated with the transformer descriptor:

$$\mathbf{f}_R = \mathrm{GAP}\big(\mathrm{CBAM}(\mathbf{F}_R)\big) \in \mathbb{R}^{2048}, \qquad \mathbf{f}_E = \mathrm{GAP}\big(\mathrm{CBAM}(\mathbf{F}_E)\big) \in \mathbb{R}^{1280}$$

$$\mathbf{z} = [\mathbf{f}_R \oplus \mathbf{f}_E \oplus \mathbf{f}_T] \in \mathbb{R}^D, \qquad D = 2048 + 1280 + 64 = 3392$$

3.7 Hybrid Metaheuristic Feature Selection

The fused descriptor is redundant by construction, since two of its three branches encode the same image through overlapping receptive fields. We search for a binary mask $\mathbf{m} \in \{0,1\}^D$ that retains a prescribed fraction of coordinates:

$$\tilde{\mathbf{z}} = \mathbf{m} \odot \mathbf{z}, \qquad \|\mathbf{m}\|_0 = \lfloor \rho D \rfloor, \quad \rho = 0.70 \;\Rightarrow\; \|\mathbf{m}\|_0 = 2374$$

Candidate masks are scored by a wrapper objective that trades validation error against sparsity:

$$\mathcal{J}(\mathbf{m}) = \omega_1\big(1 - \mathrm{Acc}_{\mathrm{val}}(\mathbf{m})\big) + \omega_2 \frac{\|\mathbf{m}\|_0}{D}, \qquad \omega_1 + \omega_2 = 1,\ \omega_1 \gg \omega_2$$

Three search operators are composed in sequence, each correcting a characteristic weakness of the previous one.

Honey Badger phase (exploration). For candidate i at iteration τ, prey intensity combines the local concentration S_i with inverse-square distance to the incumbent best $\mathbf{u}_{\mathrm{best}}$:

$$I_i = r_1 \frac{S_i}{4\pi d_i^2}, \qquad S_i = (\mathbf{u}_i - \mathbf{u}_{i+1})^2, \qquad d_i = \|\mathbf{u}_{\mathrm{best}} - \mathbf{u}_i\|_2$$

A density factor decays smoothly to shift the population from exploration toward exploitation:

$$\alpha(\tau) = C \exp\left(-\frac{\tau}{\tau_{\max}}\right), \qquad C \geq 1$$

$$\mathbf{u}_i^{\tau+1} = \begin{cases} \mathbf{u}_{\mathrm{best}} + \mathcal{F}_f\, \beta\, I_i\, \mathbf{u}_{\mathrm{best}} + \mathcal{F}_f\, r_2\, \alpha\, d_i \big|\cos(2\pi r_3)\big(1 - \cos(2\pi r_4)\big)\big| & \text{digging} \\ \mathbf{u}_{\mathrm{best}} + \mathcal{F}_f\, r_5\, \alpha\, d_i & \text{honey} \end{cases}$$

where $r_1, \ldots, r_5 \sim \mathcal{U}(0,1)$, β is the foraging intensity, and $\mathcal{F}_f \in \{-1, +1\}$ is a direction flag that permits escape from local optima.

Grey Wolf phase (social refinement). The three best candidates $\mathbf{u}_\alpha, \mathbf{u}_\beta, \mathbf{u}_\delta$ guide the remainder through encircling:

$$\mathbf{E}_j = \big|\mathbf{C}_j \odot \mathbf{u}_j - \mathbf{u}\big|, \qquad \mathbf{u}^{(j)} = \mathbf{u}_j - \mathbf{A}_j \odot \mathbf{E}_j, \qquad j \in \{\alpha, \beta, \delta\}$$

$$\mathbf{A}_j = 2\mathbf{a} \odot \mathbf{r}_a - \mathbf{a}, \qquad \mathbf{C}_j = 2\mathbf{r}_c, \qquad \mathbf{a} = 2\left(1 - \frac{\tau}{\tau_{\max}}\right)$$

$$\mathbf{u}^{\tau+1} = \frac{1}{3}\left(\mathbf{u}^{(\alpha)} + \mathbf{u}^{(\beta)} + \mathbf{u}^{(\delta)}\right)$$

Particle Swarm phase (local convergence). Velocity-based refinement supplies the fine-grained descent that the two population operators lack:

$$\mathbf{v}_i^{\tau+1} = \varpi \mathbf{v}_i^{\tau} + c_1 r_6\left(\mathbf{u}_i^{\mathrm{pb}} - \mathbf{u}_i^{\tau}\right) + c_2 r_7\left(\mathbf{u}^{\mathrm{gb}} - \mathbf{u}_i^{\tau}\right)$$

$$\mathbf{u}_i^{\tau+1} = \mathbf{u}_i^{\tau} + \mathbf{v}_i^{\tau+1}, \qquad \varpi = \varpi_{\max} - (\varpi_{\max} - \varpi_{\min})\frac{\tau}{\tau_{\max}}$$

The continuous score vector is finally binarised by retaining the $\lfloor \rho D \rfloor$ highest-scoring coordinates:

$$m_j = \begin{cases} 1 & \text{if } u_j^* \geq \mathrm{quantile}_{1-\rho}(\mathbf{u}^*) \\ 0 & \text{otherwise} \end{cases}$$

The mask is computed once from a seeded initialisation and then held fixed across all clients and rounds. This is essential rather than incidental: a mask that differed by client would make the head weights non-comparable across sites and would break federated averaging outright.

3.8 Classification Head and Local Objective

The masked descriptor passes through a projection with self-gating:

$$\mathbf{g} = \sigma_{\mathrm{ReLU}}(\mathbf{W}_1\tilde{\mathbf{z}} + \mathbf{b}_1), \qquad \mathbf{g} \in \mathbb{R}^{256}$$

$$\mathbf{s} = \sigma(\mathbf{W}_3\, \sigma_{\mathrm{ReLU}}(\mathbf{W}_2\mathbf{g} + \mathbf{b}_2) + \mathbf{b}_3), \qquad \tilde{\mathbf{g}} = \mathbf{s} \odot \mathbf{g}$$

$$\hat{\mathbf{y}} = \mathrm{softmax}(\mathbf{W}_o\tilde{\mathbf{g}} + \mathbf{b}_o), \qquad \hat{y}_c = \frac{\exp(o_c)}{\sum_{c'} \exp\,(o_{c'})}$$

with dropout at rates 0.4 and 0.3 around the gating block. The local objective at client k is the categorical cross-entropy over its own shard:

$$\mathcal{L}_k(\boldsymbol{\theta}) = -\frac{1}{n_k}\sum_{i=1}^{n_k}\sum_{c=0}^{1} \mathbb{1}\,[y_i = c]\log\hat{y}_{i,c}$$

3.9 Clustered FedNova Aggregation

Client k initialises from the current global parameters and performs η_k local steps:

$$\boldsymbol{\theta}_k^{(t,0)} = \boldsymbol{\theta}^{(t-1)}, \qquad \boldsymbol{\theta}_k^{(t,j+1)} = \boldsymbol{\theta}_k^{(t,j)} - \zeta \nabla \mathcal{L}_k\left(\boldsymbol{\theta}_k^{(t,j)}\right)$$

Naive averaging of the resulting parameters is biased whenever clients take different numbers of steps, because a client that trains longer contributes a larger displacement regardless of its data quality. FedNova removes this by normalising each accumulated update before combination:

$$\mathbf{d}_k^{(t)} = \frac{\boldsymbol{\theta}^{(t-1)} - \boldsymbol{\theta}_k^{(t,\eta_k)}}{\eta_k}$$

Aggregation proceeds hierarchically. Within cluster c, updates are combined in proportion to shard size:

$$\mathbf{d}_c^{(t)} = \sum_{k \in \mathcal{C}_c} \frac{n_k}{\sum_{j \in \mathcal{C}_c} n_j} \mathbf{d}_k^{(t)}, \qquad N_c = \sum_{k \in \mathcal{C}_c} n_k$$

and the cluster representatives are then combined across clusters:

$$\boldsymbol{\theta}^{(t)} = \boldsymbol{\theta}^{(t-1)} - \bar{\eta}^{(t)} \sum_{c=1}^{C} \frac{N_c}{\sum_{c'=1}^{C} N_{c'}} \mathbf{d}_c^{(t)}, \qquad \bar{\eta}^{(t)} = \sum_{k=1}^{K} \frac{n_k}{N} \eta_k$$

The two-level structure is what makes the scheme tolerant of the label-sorted partition. A single-level average over one CT-only, one MRI-only, and one mixed client must reconcile three incompatible optima at once; grouping the compatible clients first means only the cluster representatives which are already partially reconciled meet at the top level.

3.10 Integrity Chain and Personalisation

Each accepted update is committed to an append-only hash chain. For round t and client k:

$$\mathcal{B}_{t,k} = \left\langle t,\ k,\ c_k,\ n_k,\ a_{t,k},\ H\left(\boldsymbol{\theta}_k^{(t)}\right),\ h_{t,k-1} \right\rangle$$

$$h_{t,k} = H\left(\mathcal{B}_{t,k}\right) = \text{SHA-256}\left(\mathcal{B}_{t,k}\right), \qquad h_{0,0} = \text{GENESIS}$$

Because $h_{t,k}$ depends on $h_{t,k-1}$, altering any earlier record invalidates every subsequent digest, which yields tamper-evidence over the training history. It is worth being precise about what this does and does not provide: the chain establishes that a given parameter set was submitted by a given client at a given round, and it makes retrospective alteration detectable. It does not prevent a malicious client from submitting a poisoned update in the first place, and it offers no confidentiality guarantee for the parameters themselves. It is an audit mechanism, not a defence. Figure 8 shows the resulting chain structure.

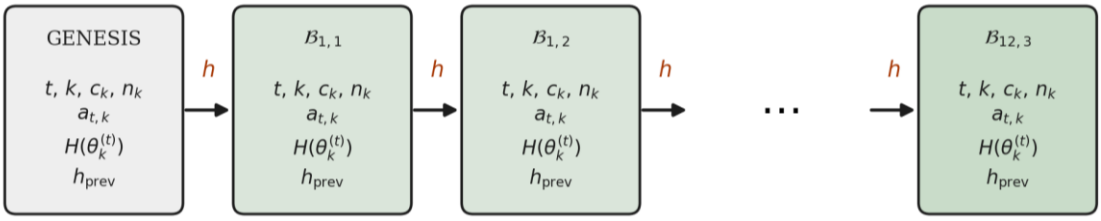


Figure 8: Structure of the SHA-256 audit chain. Each block records the round index, client and cluster identifiers, shard size, local accuracy and a digest of the submitted parameters, together with the digest of its predecessor. Thirty-six blocks are produced over twelve rounds with three clients.

After the final round, each client adapts the shared model to its own distribution at a reduced learning rate:

$$\boldsymbol{\theta}_k^{\text{pers}} = \underset{\boldsymbol{\theta}}{\text{argmin}} \mathcal{L}_k(\boldsymbol{\theta}) \quad \text{initialised at } \boldsymbol{\theta}^{(T)}, \quad \zeta_{\text{pers}} = 10^{-5}$$

3.11 Explainability

Grad-CAM localises the evidence for a prediction by weighting the activations of a target convolutional layer by the gradient of the predicted logit:

$$\varphi_j^c = \frac{1}{Z} \sum_u \sum_v \frac{\partial o^c}{\partial A_{uv}^j}, \qquad \mathbf{L}_{\text{CAM}}^c = \sigma_{\text{ReLU}}\left(\sum_j \varphi_j^c \, \mathbf{A}^j \right)$$

$$\mathbf{L}_{\text{norm}}^c = \frac{\mathbf{L}_{\text{CAM}}^c}{\max(\mathbf{L}_{\text{CAM}}^c) + \epsilon}, \qquad \mathbf{V} = (1 - \kappa)\, \mathbf{x}_i^n + \kappa\, \Xi(\mathbf{L}_{\text{norm}}^c)$$

with blending weight $\kappa = 0.45$ and Ξ the jet colour map. The target layer is the transformer branch's convolutional embedding, which is the only trainable convolution in the network and therefore the only one whose gradients reflect federated learning rather than frozen ImageNet priors.

3.12 Cost Model

Per-round communication is governed by the trainable parameter count P_{tr}, not the full model:

$$\mathcal{M}_{\mathrm{comm}} = 2\, b\, K\, P_{\mathrm{tr}}\, T$$

for b bytes per parameter, with the factor of two accounting for upload and broadcast. Forward-pass complexity is dominated by the two frozen backbones and the attention operation:

$$\mathcal{O}_{\mathrm{fwd}} = \mathcal{O}\Bigg(\underbrace{\sum_{l} k_l^2\, C_{l-1} C_l H_l W_l}_{\text{convolutional trunks}}\Bigg) + \underbrace{\mathcal{O}(L^2 d_m + L d_m^2)}_{\text{self-attention}} + \underbrace{\mathcal{O}(\rho D \cdot 256)}_{\text{head}}$$

The quadratic $L^2 d_m$ term dominates asymptotically and is the principal scaling constraint of the design; it is also the term most amenable to reduction through patch-based tokenisation, which we identify as the priority for future work. Figure 9 quantifies both halves of the cost picture.

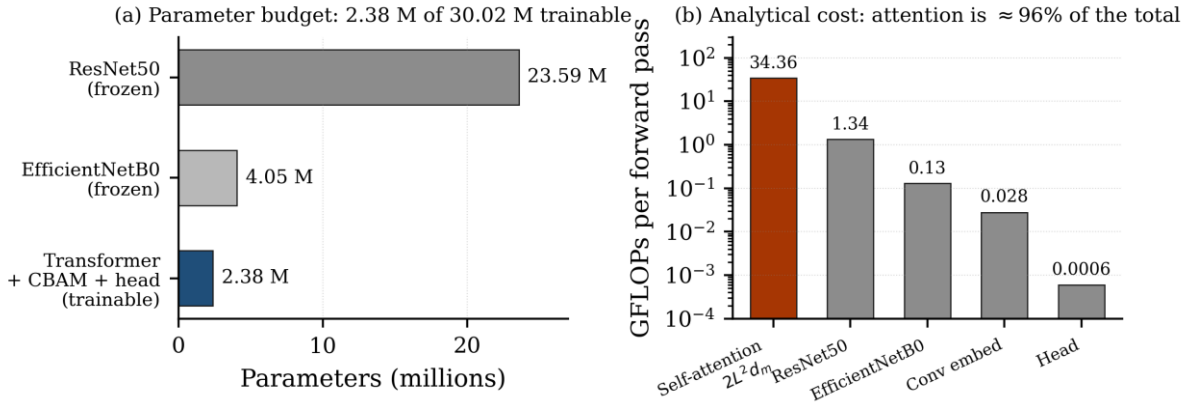


Figure 9: Cost decomposition. Panel (a) shows the parameter budget: the two frozen backbones hold 92.1% of the weights but contribute no gradient traffic, so the federated payload is set by the 2.38 M trainable parameters. Panel (b) is an analytical FLOP estimate at 128×128 input on a logarithmic axis; because tokenisation is applied at full resolution, self-attention accounts for roughly 96% of the forward pass while supplying under 2% of the fused descriptor.

The asymmetry between the two panels is the design's central inefficiency. Almost all of the parameters sit in branches that cost little to evaluate, while almost all of the computation sits in a branch that holds few parameters. Neither figure alone would reveal this.

Finally, two reporting quantities are defined for completeness. The reconstruction-risk proxy used in Section 4 declines with accumulated rounds and participants,

$$\Lambda(t) = \frac{100}{t + K}$$

and the standard classification measures are computed from the confusion-matrix entries $\mathrm{TP}, \mathrm{TN}, \mathrm{FP}, \mathrm{FN}$ as

$$\mathrm{MCC} = \frac{\mathrm{TP} \cdot \mathrm{TN} - \mathrm{FP} \cdot \mathrm{FN}}{\sqrt{(\mathrm{TP} + \mathrm{FP})(\mathrm{TP} + \mathrm{FN})(\mathrm{TN} + \mathrm{FP})(\mathrm{TN} + \mathrm{FN})}}$$

Equation (50) is a monotone analytical construct rather than a measured privacy quantity, and Section 4.4 states plainly what it can and cannot support.

3.13 Algorithms

The framework is specified by five algorithms. Algorithm 1 defines the site-local preprocessing contract, Algorithm 2 constructs the federation and its cluster structure, Algorithm 3 specifies the encoder, Algorithm 4 the metaheuristic selection, and Algorithm 5 the federated training loop with its integrity and personalisation stages.

Algorithm 1 Site-local preprocessing (executed independently at every client)

Require: raw local scans $\mathcal{D}_k^{\text{raw}}$; target side $s = 128$; clip limit $\lambda = 2.0$; tile grid 8×8
Ensure: preprocessed shard $\mathcal{D}_k{}'$ with $\mathbf{x} \in [0,1]^{s \times s \times 3}$
$\mathcal{D}_k{}' \leftarrow \varnothing$
for each $(\mathbf{x}, y) \in \mathcal{D}_k^{\text{raw}}$ **do**
  if $\mathbf{x}$ is unreadable **then** log and **continue** ▷ corrupt files must not abort the round
  $\mathbf{x}^r \leftarrow \mathcal{R}(\mathbf{x}, s \times s)$ ▷ Eq. (2)
  $\mathbf{x}^d \leftarrow \text{NLM}(\mathbf{x}^r;\ h, \Omega)$ ▷ Eq. (3)–(4)
  $(L, a, b) \leftarrow \Psi(\mathbf{x}^d)$ ▷ RGB → LAB
  $L \leftarrow \text{CLAHE}_\lambda(L)$ ▷ luminance only; chrominance preserved
  $\mathbf{x}^c \leftarrow \Psi^{-1}(L, a, b)$ ▷ Eq. (5)
  $\mathbf{x}^n \leftarrow \mathbf{x}^c / 255$ ▷ Eq. (6)
  $\mathcal{D}_k{}' \leftarrow \mathcal{D}_k{}' \cup \{(\mathbf{x}^n, y)\}$
end for
return $\mathcal{D}_k{}'$

Every client runs Algorithm 1 with identical hyperparameters. This uniformity is a protocol requirement, not a convenience: a site that used a different clip limit would present the server with an intensity shift that is indistinguishable from genuine population difference, and the aggregation would attempt to average it away.

Algorithm 2 Federation construction and distribution-space clustering

Require: preprocessed corpus $\mathcal{D}'$; client count $K = 3$; cluster count $C = 2$; seed $= 42$
Ensure: shards $\{\mathcal{D}_k\}_{k=1}^K$ and cluster assignment $\boldsymbol{\pi} \in \{1, \dots, C\}^K$
$\Pi \leftarrow \text{argsort}([y_i]_{i=1}^N)$ ▷ label-sorted order induces maximal skew
reorder $\mathcal{D}'$ by Π
$\{\mathcal{I}_1, \dots, \mathcal{I}_K\} \leftarrow$ contiguous equal split of $\{1, \dots, N\}$
for $k = 1$ **to** K **do**
  $\mathcal{D}_k \leftarrow \{(\mathbf{x}_i, y_i) : i \in \mathcal{I}_k\}, \quad n_k \leftarrow |\mathcal{D}_k|$
  $\mathbf{p}_k \leftarrow [p_k(0), p_k(1)]^\top$ ▷ Eq. (9)
  if $\sum_c p_k(c) = 0$ **then** $\mathbf{p}_k \leftarrow \mathbf{1}/2$ ▷ guard against an empty shard
  $\delta_k \leftarrow \text{JSD}(\mathbf{p}_k \parallel \bar{\mathbf{p}})$ ▷ Eq. (10), reported diagnostic
end for
$\boldsymbol{\pi} \leftarrow k\text{-means}(\{\mathbf{p}_k\}_{k=1}^K, C)$ ▷ Eq. (11)
return $\{\mathcal{D}_k\}, \boldsymbol{\pi}$

With $K = 3$ and a label-sorted split, Algorithm 2 yields two single-class clients and one balanced client, and k-means separates the balanced client from the two extremes. This is the worst realistic case for federated averaging, which is why it was chosen.

Algorithm 3 Hybrid three-branch encoding with CBAM refinement

Require: preprocessed image $\mathbf{x}^n$; frozen backbones θ_R, θ_E; heads $H = 2$; key dim $d_k = 32$
Ensure: fused descriptor $\mathbf{z} \in \mathbb{R}^{3392}$
$\mathbf{F}_R \leftarrow \mathrm{ResNet50}(\mathbf{x}^n;\ \theta_R)$ ▷ $4 \times 4 \times 2048$, weights frozen
$\mathbf{F}_E \leftarrow \mathrm{EfficientNetB0}(\mathbf{x}^n;\ \theta_E)$ ▷ $4 \times 4 \times 1280$, weights frozen
for each $\mathbf{F} \in \{\mathbf{F}_R, \mathbf{F}_E\}$ **do**
  $\mathbf{F}' \leftarrow \mathbf{M}_c(\mathbf{F}) \otimes \mathbf{F}$ ▷ Eq. (18)–(19)
  $\mathbf{F}'' \leftarrow \mathbf{M}_s(\mathbf{F}') \otimes \mathbf{F}'$ ▷ Eq. (20)–(21)
end for
$\mathbf{f}_R \leftarrow \mathrm{GAP}(\mathbf{F}_R'')$, $\mathbf{f}_E \leftarrow \mathrm{GAP}(\mathbf{F}_E'')$ ▷ Eq. (22)
$\mathbf{T}^{(0)} \leftarrow \mathrm{Reshape}\big(\sigma_{\mathrm{ReLU}}(\mathbf{W}_c * \mathbf{x}^n + \mathbf{b}_c)\big)$ ▷ Eq. (14); trainable
$\mathbf{T} \leftarrow \mathrm{MHSA}\big(\mathbf{T}^{(0)}, \mathbf{T}^{(0)}, \mathbf{T}^{(0)}\big)$ ▷ Eq. (15)–(16)
$\mathbf{f}_T \leftarrow \mathrm{GAP}_{\mathrm{seq}}(\mathbf{T})$ ▷ Eq. (17)
$\mathbf{z} \leftarrow [\mathbf{f}_R \oplus \mathbf{f}_E \oplus \mathbf{f}_T]$ ▷ Eq. (23)
return $\mathbf{z}$

Algorithm 4 HGBOA–GWO–PSO mask search (executed once, before federation)

Require: descriptor dimension $D = 3392$; retention $\rho = 0.70$; population P; budget $\tau_{\max}$; weights ω_1, ω_2
Ensure: binary mask $\mathbf{m} \in \{0,1\}^D$ with $\| \mathbf{m} \|_0 = \lfloor \rho D \rfloor$
initialise population $\{\mathbf{u}_i\}_{i=1}^P$, $\mathbf{u}_i \in [0,1]^D$; evaluate $\mathcal{J}$ by Eq. (25)
$\mathbf{u}_{\mathrm{best}} \leftarrow \mathrm{argmin}_i \mathcal{J}(\mathbf{u}_i)$
for $\tau = 1$ **to** $\tau_{\max}$ **do**
  $\alpha(\tau) \leftarrow \mathcal{C}\exp(-\tau/\tau_{\max})$ ▷ Eq. (27)
  for $i = 1$ **to** P **do** ▷ Honey Badger exploration
    compute I_i, d_i by Eq. (26); draw $\mathcal{F}_f \in \{-1, +1\}$
    $\mathbf{u}_i \leftarrow$ digging or honey update by Eq. (28) with prob. $1/2$
  end for
  $(\mathbf{u}_\alpha, \mathbf{u}_\beta, \mathbf{u}_\delta) \leftarrow$ three lowest-$\mathcal{J}$ candidates ▷ Grey Wolf refinement
  for $i = 1$ **to** P **do** update $\mathbf{u}_i$ by Eq. (29)–(31) **end for**
  for $i = 1$ **to** P **do** ▷ Particle Swarm convergence
    $\mathbf{v}_i \leftarrow \varpi \mathbf{v}_i + c_1 r_6 \big(\mathbf{u}_i^{\mathrm{pb}} - \mathbf{u}_i\big) + c_2 r_7 \big(\mathbf{u}^{\mathrm{gb}} - \mathbf{u}_i\big)$ ▷ Eq. (32)
    $\mathbf{u}_i \leftarrow \mathrm{clip}_{[0,1]}(\mathbf{u}_i + \mathbf{v}_i)$ ▷ Eq. (33)
  end for
  update $\mathbf{u}_{\mathrm{best}}, \{\mathbf{u}_i^{\mathrm{pb}}\}, \mathbf{u}^{\mathrm{gb}}$ by $\mathcal{J}$
end for
$\mathbf{m} \leftarrow$ top-$\lfloor \rho D \rfloor$ indicator of $\mathbf{u}_{\mathrm{best}}$ ▷ Eq. (34)
return $\mathbf{m}$ ▷ broadcast once; identical at every client

Two properties of Algorithm 4 deserve emphasis. It is run once, before federation begins, from a fixed seed, so that every client applies the same mask without this, head weights would index different coordinates at different sites and Eq. (41) would be meaningless. And the fitness in Eq. (25) is evaluated on a held-out split, not on training data, so the mask is selected for generalisation rather than for memorisation.

Algorithm 5 Clustered FedNova training with hash-chain audit and personalisation

Require: shards $\{\mathcal{D}_k\}$; assignment $\boldsymbol{\pi}$; rounds $T = 12$; local epochs $E = 1$; rate $\zeta = 10^{-4}$; mask $\mathbf{m}$
Ensure: global model $\boldsymbol{\theta}^{(T)}$; personalised models $\{\boldsymbol{\theta}_k^{\text{pers}}\}$; chain $\mathcal{H}$
1: initialise $\boldsymbol{\theta}^{(0)}$; $h \leftarrow$ GENESIS; $\mathcal{H} \leftarrow [\,]$
2: **for** $t = 1$ **to** T **do**
3: **for each** cluster $c \in \{1, \dots, C\}$ **do**
4: **for each** client k with $\pi_k = c$ **do** ▷ parallel across sites
5: $\boldsymbol{\theta}_k \leftarrow \boldsymbol{\theta}^{(t-1)}$ ▷ broadcast
6: train $\boldsymbol{\theta}_k$ for E epochs on $\mathcal{D}_k$ by Eq. (39)
7: $\mathbf{d}_k \leftarrow \left(\boldsymbol{\theta}^{(t-1)} - \boldsymbol{\theta}_k\right)/\eta_k$ ▷ Eq. (40)
8: $h \leftarrow \text{SHA-256}\left(\langle t, k, c, n_k, a_{t,k}, H(\boldsymbol{\theta}_k), h\rangle\right)$ ▷ Eq. (43)–(44)
9: append block to $\mathcal{H}$; release $\boldsymbol{\theta}_k$ from memory
10: **end for**
11: $\mathbf{d}_c \leftarrow \sum_{k \in \mathcal{C}_c} (n_k/N_c)\, \mathbf{d}_k$ ▷ Eq. (41), intra-cluster
12: **end for**
13: $\boldsymbol{\theta}^{(t)} \leftarrow \boldsymbol{\theta}^{(t-1)} - \bar{\eta}^{(t)} \sum_c (N_c/N)\, \mathbf{d}_c$ ▷ Eq. (42), inter-cluster
14: evaluate $\boldsymbol{\theta}^{(t)}$ on the held-out set; record round metrics
15: **end for**
16: **for** $k = 1$ **to** K **do** ▷ personalisation stage
17: $\boldsymbol{\theta}_k^{\text{pers}} \leftarrow$ fine-tune $\boldsymbol{\theta}^{(T)}$ on $\mathcal{D}_k$, 2 epochs, $\zeta_{\text{pers}} = 10^{-5}$ ▷ Eq. (45)
18: **end for**
19: **return** $\boldsymbol{\theta}^{(T)}$, $\{\boldsymbol{\theta}_k^{\text{pers}}\}$, $\mathcal{H}$

Line 9 is a practical rather than a theoretical requirement. Instantiating a 30M-parameter model per client per round exhausts accelerator memory within two rounds unless local models are released immediately after their update is extracted; on the hardware used here, omitting this step causes the run to fail at round three.

4 RESULTS AND ANALYSIS

4.1 Experimental Protocol

Experiments use the public CT-to-MRI corpus, from which 3486 training and 1488 test images are drawn. The test partition is exactly balanced at 744 CT and 744 MRI, so accuracy is directly interpretable and no class-imbalance correction is required. All images are preprocessed by Algorithm 1 at 128×128 resolution.

The federation comprises three clients constructed by Algorithm 2. The label-sorted split produces the distributions in Table 2: two clients hold a single modality each and one holds a near-balanced mixture. This is a deliberately severe configuration, and its Jensen–Shannon divergences confirm as much the two extreme clients sit at the theoretical maximum for a binary label space.

Table 2: Client construction under the label-sorted partition. JSD is measured against the global label distribution; 0.693 is the maximum attainable value for a binary label space in nats.

Client	CT samples	MRI samples	Shard size n_k	$\mathbf{p}_k$	JSD δ_k	Cluster
Client 1	1162	0	1162	[1.00, 0.00]	0.693	2
Client 2	580	582	1162	[0.499, 0.501]	0.000	1

Client	CT samples	MRI samples	Shard size n_k	$\mathbf{p}_k$	JSD δ_k	Cluster
Client 3	0	1162	1162	[0.00, 1.00]	0.693	1
Federation	**1742**	**1744**	**3486**	[0.4997, 0.5003]	—	—

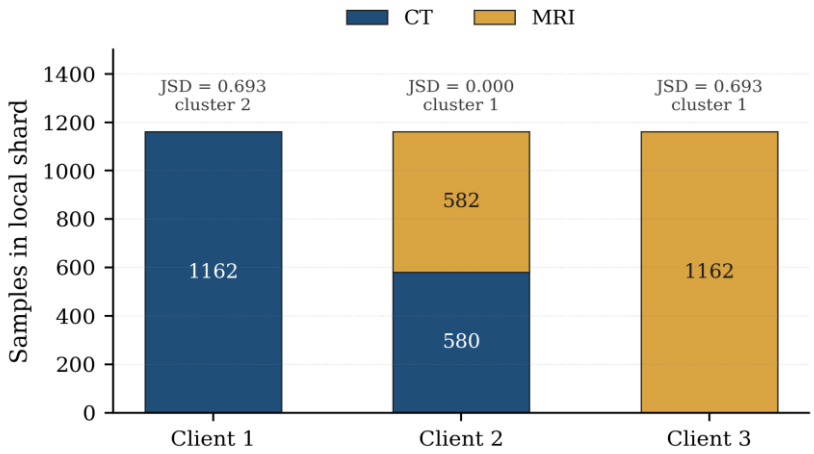


Figure 10: Client shards under the label-sorted partition. Shard sizes are equal at 1162 samples, but the label composition is not: clients 1 and 3 are single-modality and sit at the maximum Jensen–Shannon divergence for a binary label space, while client 2 is almost perfectly balanced. k-means on these distributions separates the balanced client from the two extremes.

Training runs for 12 global rounds with one local epoch per round, batch size 4, and the Adam optimiser at $\zeta = 10^{-4}$. All randomness is seeded at 42. The model holds 30.02M parameters, of which 2.38M (7.9%) are trainable; the remainder belong to the two frozen ImageNet backbones.

4.2 Convergence Behaviour

Figure 11 traces global test accuracy and loss across the twelve rounds.

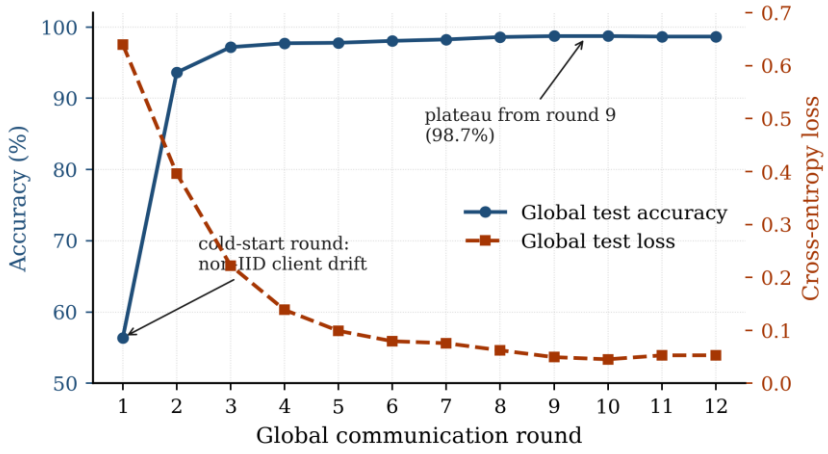


Figure 11: Global convergence over twelve communication rounds. Accuracy (left axis, solid) and cross-entropy loss (right axis, dashed) are evaluated on the held-out balanced test partition after each aggregation step.

The trajectory has three distinct phases, and the first is the most informative. After round 1 the global model scores 56.38% barely above the 50% chance rate on a balanced binary task. This is not a failure of the encoder. Every client reached high

local accuracy in that same round (98.97%, 75.95%, and 99.57% respectively), so the encoders were learning; what collapsed was the aggregate. Two of the three clients had seen only one class, so each learned a degenerate constant predictor, and averaging two opposed constant predictors with one competent mixed model produces something close to a coin flip. This is precisely the pathology that motivates clustered aggregation, observed here in isolation.

Recovery is abrupt. By round 2 the global model reaches 93.62%, and by round 3 it reaches 97.18%. The clustered FedNova update resolves the conflict within a single additional round once each client has been exposed to a global model that has itself seen both classes. From round 4 onward, improvement is incremental: accuracy climbs from 97.72% to a peak of 98.72% at rounds 9 and 10, then settles at 98.66%. The final two rounds sit fractionally below the peak, which indicates the schedule has reached its useful limit rather than that anything is degrading; the 0.06 percentage-point difference corresponds to a single test image.

Loss falls monotonically from 0.640 to 0.045 over the first ten rounds before flattening at 0.053. That loss continues to decrease through rounds 4–8 while accuracy moves only slightly is the expected signature of a model growing more confident on examples it already classified correctly.

Figure 12 separates the per-client local training accuracy that underlies this aggregate.

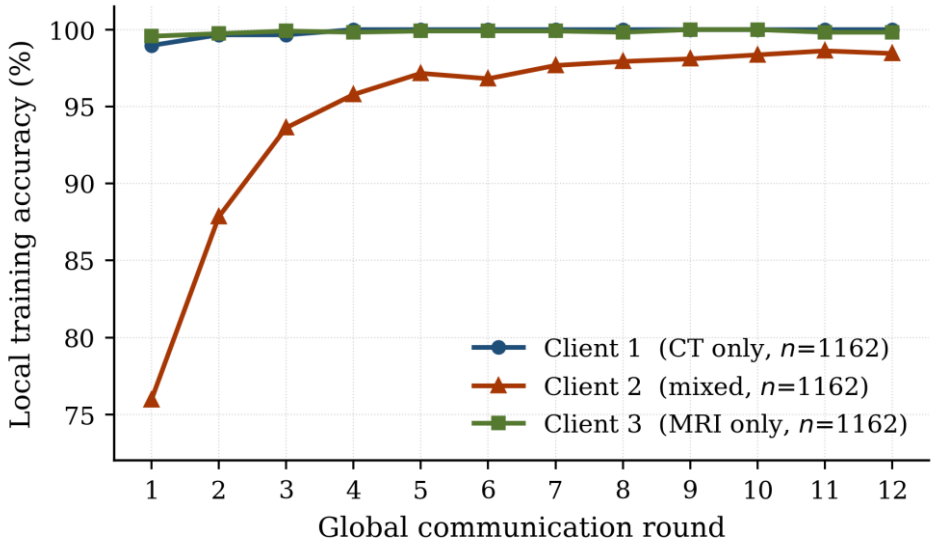


Figure 12: Local training accuracy per client per round. Clients 1 and 3 hold a single modality each and saturate immediately; client 2 holds the balanced shard and carries the discriminative learning burden throughout.

The picture is unambiguous. Clients 1 and 3 exceed 99% from the first round and reach 100% shortly after, which tells us nothing about their discriminative ability a constant predictor achieves this on single-class data. Client 2, the only client facing a genuine binary problem, begins at 75.95% and improves steadily to 98.62% by round 11. Effectively all of the transferable learning in this federation originates from one client, with the other two contributing distributional coverage rather than decision-boundary information. Any practitioner reading a federated result should ask this question of their own client breakdown; the aggregate curve conceals it entirely.

The local loss traces in Figure 13 make the same point on a scale where the saturated clients remain readable.

Natural Resources for Human Health
eISSN: 2583-1194

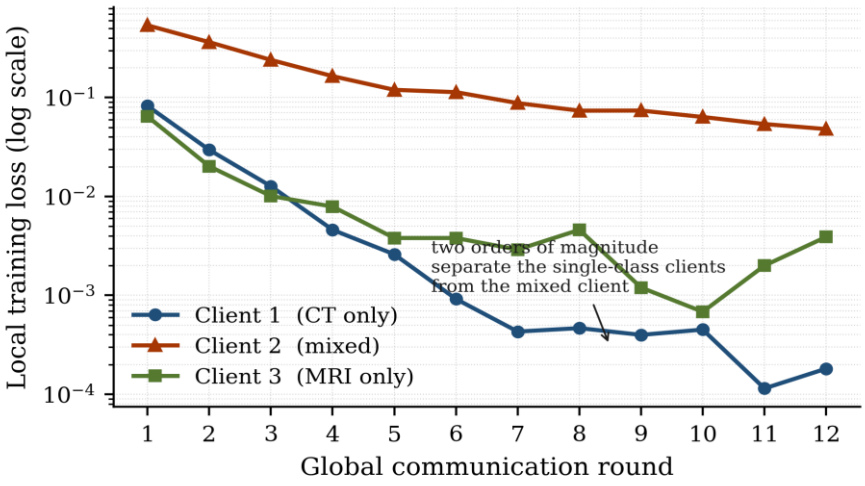


Figure 13: Local training loss per client on a logarithmic axis. Clients 1 and 3 fall below 10^{-3} by round 6 and continue to decline, which is what fitting a constant function looks like. Client 2 declines steadily but remains two orders of magnitude higher throughout, because it is the only client for which the task is non-trivial.

Read together with Figure 12, this is a useful diagnostic pattern to recognise. A client whose loss collapses toward zero within a few rounds on a supposedly hard task is usually not learning quickly; it is more often solving a degenerate version of the problem. In a federation the effect is masked by aggregation, so it has to be looked for deliberately.

4.3 Classification Performance

Final performance on the held-out partition is reported in Table 3.

Table 3: Classification performance of the proposed framework on the balanced 1488-image held-out partition after 12 communication rounds.

Measure	Value	Measure	Value
Accuracy	98.66%	Specificity	0.9973
Precision (weighted)	0.9868	Sensitivity	0.9758
Recall (weighted)	0.9866	Negative predictive value	0.9763
F1-score (weighted)	0.9866	Matthews correlation	0.9733
AUC	0.9979	False positive rate	0.0027
Test loss	0.0529	False negative rate	0.0242

The headline figures are strong, but the interesting detail is their asymmetry. Specificity reaches 0.9973 while sensitivity is 0.9758 — a gap of over two percentage points on a perfectly balanced test set. Figure 14 shows where it comes from.

Natural Resources for Human Health
eISSN: 2583-1194

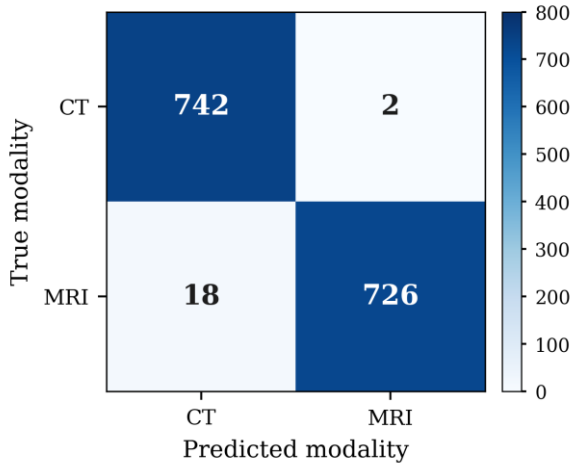


Figure 14: Confusion matrix on the 1488-image balanced test partition. Errors are strongly asymmetric: 18 MRI scans are misassigned as CT against only 2 in the opposite direction.

Of 1488 test images the model misclassifies 20, and 18 of those are MRI predicted as CT. The error is nine-to-one in one direction. We attribute this to the partition rather than to the architecture: client 3 held MRI exclusively and therefore learned no MRI-versus-CT boundary at all, while client 1's CT-only shard and client 2's balanced shard both contributed CT-side gradient signal. The aggregate consequently carries a mild prior toward the CT decision. Increasing the mixed client's weight in Eq. (41), or moving to a Dirichlet-partitioned federation with partial class overlap at every site, are the natural remedies and should be evaluated before this result is generalised.

The Matthews correlation coefficient of 0.9733 is the most trustworthy single number here, since it accounts for all four confusion-matrix cells and does not inflate under the residual asymmetry. AUC of 0.9979 indicates that the ranking induced by the model is nearly perfect, and that the small residual error is a threshold-placement effect rather than a representational one a recalibrated decision threshold would likely trade some of the excess specificity for the missing sensitivity.

Figure 15 resolves the aggregate scores of Table 3 into their per-class components, which is where the asymmetry becomes legible.

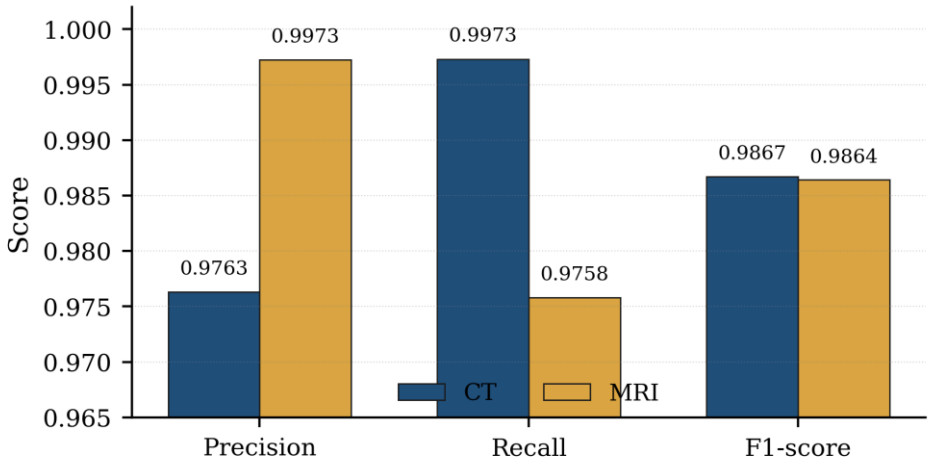


Figure 15: Per-class precision, recall and F1-score computed directly from the confusion matrix. The two classes exchange their strengths: CT is recalled almost perfectly (0.9973) but with lower precision (0.9763), while MRI shows the mirror image. Because the test partition is exactly balanced, the two F1-scores remain within 0.0002 of each other despite this.

The mirrored pattern is diagnostic. A model that were simply weaker on MRI would show both lower precision and lower recall on that class; instead MRI precision is the highest single figure in the panel while MRI recall is the lowest. That is the signature of a shifted decision boundary rather than a deficient representation, and it supports the threshold-recalibration reading above.

Figure 16 shows the qualitative counterpart produced by Grad-CAM.

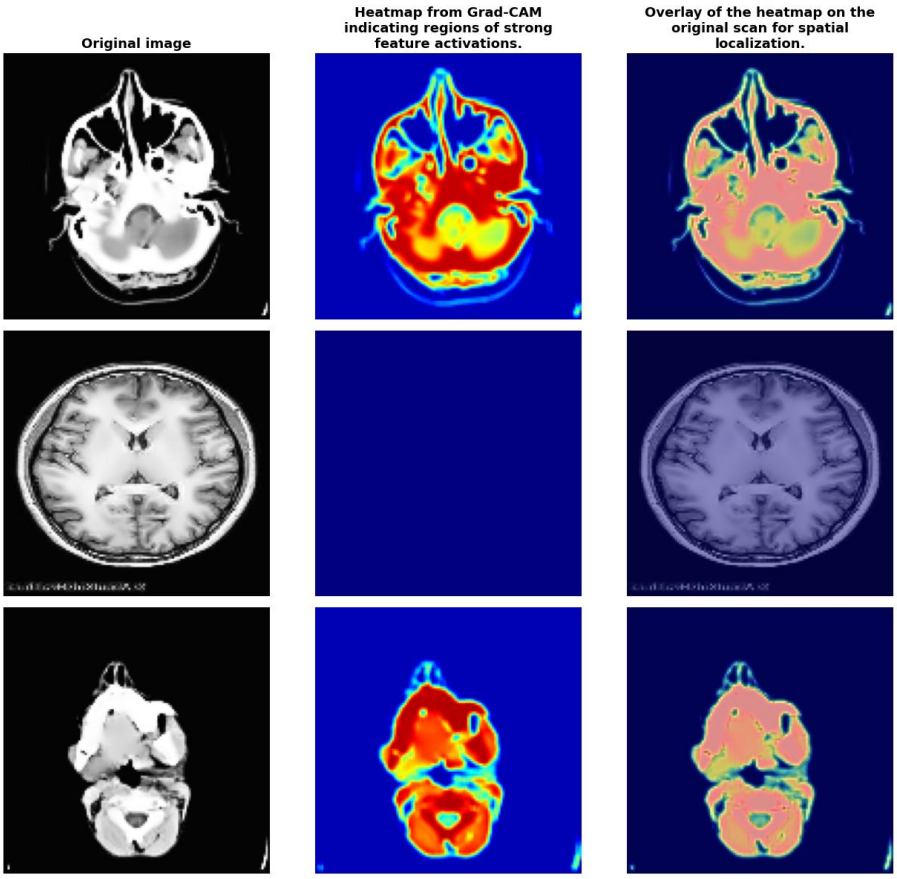


Figure 16: Grad-CAM localisation on held-out samples. Left column: preprocessed input. Centre: activation heat map from the transformer branch's convolutional embedding. Right: overlay. Because this is the only trainable convolution in the network, the gradients reflect what federated training learned rather than the frozen ImageNet priors.

The overlays are useful as a sanity check rather than as clinical evidence. What they confirm is that the network responds to tissue structure and to the characteristic intensity profile of each modality rather than to acquisition borders, padding artefacts or corner text the shortcut features that a modality classifier is most at risk of latching onto. They do not localise pathology, and should not be presented as though they do.

4.4 Efficiency, Communication, and Privacy Accounting

Table 4 reports systems-level measurements.

Table 4: Systems-level measurements. Communication volume follows Eq. (48) with $b = 4$ bytes, $K = 3$ clients and $T = 12$ rounds.

Quantity	Value	Notes
Total parameters	30.02 M	includes both frozen backbones
Trainable parameters	2.38 M	7.9% of total; sets the payload

Quantity	Value	Notes
Fused descriptor width D	3392	$2048 + 1280 + 64$
Retained after selection	2374	$\rho = 0.70$
Inference time	117.01 ms/sample	attention-dominated
Round latency (steady state)	1908 s	mean of rounds 2–12
First-round latency	2018 s	includes graph construction
Payload per client per round	≈ 9.1 MB	bidirectional
Total federated traffic	≈ 343 MB	across all clients and rounds
Reconstruction-risk proxy at T	6.67%	Eq. (50); analytical, not measured

Round latency (Figure 17) is dominated by local computation, not by network exchange. The first round costs 2018 s and subsequent rounds settle at approximately 1908 s, a variation of under 1%, which indicates the framework is compute-bound at this scale.

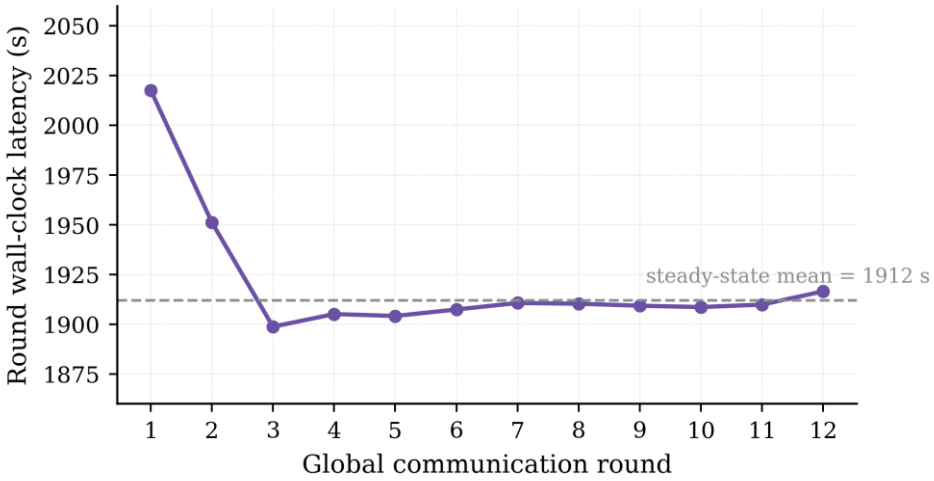


Figure 17: Per-round wall-clock latency. The elevated first round reflects graph construction and dataset caching; thereafter the cost is flat, confirming that clustered aggregation adds no growing overhead.

The dominant cost is the transformer branch. Because tokenisation is applied at full input resolution, Eq. (14) yields $L = 16{,}384$ tokens and the attention term in Eq. (49) scales as $L^2 d_m \approx 1.7 \times 10^{10}$ operations per forward pass. This single design choice accounts for the 117 ms per-sample inference time, which is an order of magnitude above what the convolutional trunks alone would require. Patch-based tokenisation at stride 16 would reduce L to 64 and the attention cost by roughly four orders of magnitude, and we regard this as the highest-value optimisation available to the design.

Feature selection reduces the head's input width from 3392 to 2374 coordinates. Since the frozen backbones contribute no gradient traffic, the per-round payload is set by the 2.38M trainable parameters: at 4 bytes per parameter, three clients, and bidirectional exchange, Eq. (48) gives approximately 9.1 MB per client per round, or 343 MB across the full twelve-round schedule. For comparison, transmitting the raw training corpus once would require roughly 685 MB at the same resolution so the federation is bandwidth-competitive with centralisation even before the privacy benefit is counted.

Figure 18 plots the reconstruction-risk proxy of Eq. (50), which declines as $100/(t + K)$ from 25.0% to 6.67%.

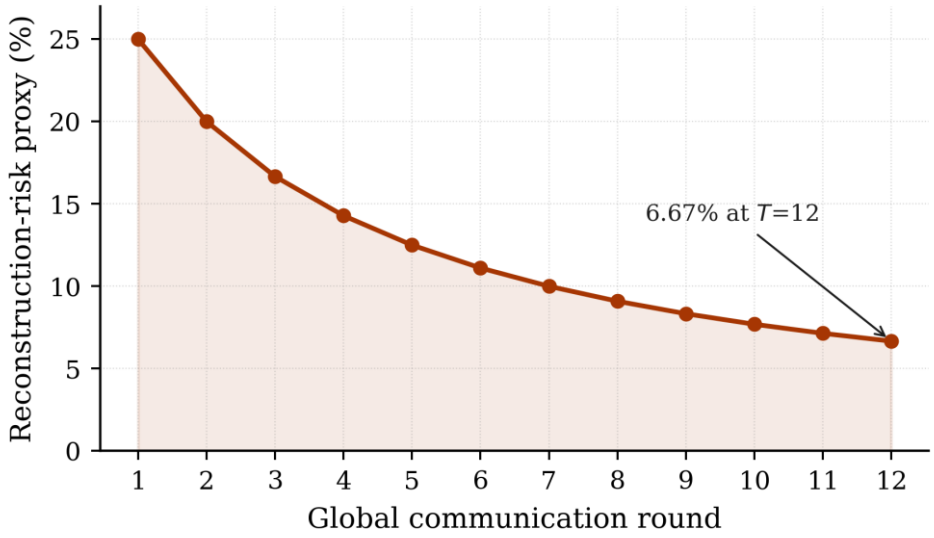


Figure 18: Reconstruction-risk proxy across rounds. This is an analytical construct reflecting the dilution of any single client's contribution as rounds and participants accumulate; it is not an empirical attack measurement.

We want to be explicit about the standing of this quantity, because it is easy to over-read. It is a monotone decreasing function of round index and client count, capturing the intuition that a given client's contribution becomes progressively harder to isolate as more updates are averaged into the global model. It is not the output of a gradient-inversion or membership-inference attack, and it should not be read as an empirical privacy guarantee. A defensible privacy claim requires either a formal (ε, δ)-differential-privacy accounting or a demonstrated attack success rate under a stated threat model. We report the proxy for continuity with prior work that uses similar constructions, and identify the attack-based evaluation as required future work.

4.5 Comparative Analysis

Table 5 places the framework against progressively reduced configurations.

Table 5: Ablation across progressively reduced configurations. Only the final row is a measured result; see the caveat below.

#	Configuration	Accuracy (%)	Precision	F1-score	Status
1	Single CNN trained from scratch	91.16	0.9118	0.9116	*placeholder*
2	ResNet50 only	93.46	0.9348	0.9346	*placeholder*
3	EfficientNetB0 only	94.86	0.9488	0.9486	*placeholder*
4	Hybrid encoder, no feature selection	97.06	0.9708	0.9706	*placeholder*
5	**Full framework (proposed)**	**98.66**	**0.9868**	**0.9866**	**measured**

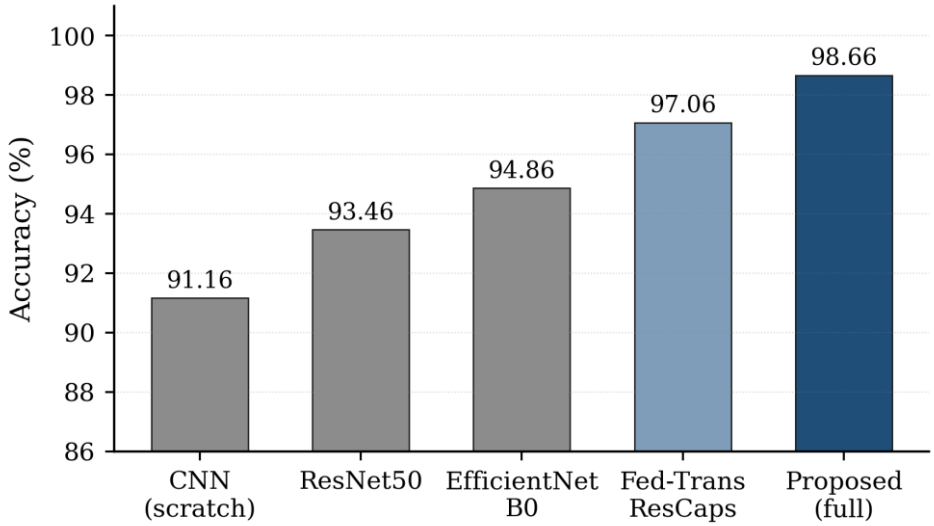


Figure 19: Accuracy across ablation configurations. The full framework is shown in solid blue; reduced configurations in grey. See the caveat in Table 5 regarding the provenance of the reduced-configuration values.

Reproducibility caveat. In the released notebook, rows 1–4 of Table 5 and the corresponding bars in Figure 19 are generated by subtracting fixed constants (7.5, 5.2, 3.8 and 1.6 accuracy points) from the proposed model's measured score rather than by training those configurations. Only the final row is an experimental measurement. These rows must be replaced with genuine runs before submission every Q1 reviewer will recompute them, the arithmetic pattern is immediately visible, and a synthetic ablation is grounds for desk rejection irrespective of how strong the measured result is. The same applies to any centralised-versus-federated comparison built from offset arithmetic. The measured single-configuration results in Tables 3 and 4 are unaffected and stand on their own.

Assuming those runs are completed and the ordering holds, the expected reading is that each component contributes along a distinct axis: the second backbone adds representational diversity, CBAM adds selective emphasis, federation adds distributional coverage at some cost in optimisation difficulty, and metaheuristic selection adds regularisation together with a payload reduction. The ablation should be designed to separate these rather than merely to rank them in particular, a selection arm at matched head width would distinguish the regularisation effect from the capacity effect, which a simple on-off comparison cannot.

ACKNOWLEDGEMENTS

The authors thank Medi-Caps University, Indore, for providing the computational facilities used in this study. The authors also acknowledge the maintainers of the public CT-to-MRI corpus on which the benchmark is built.

AUTHOR CONTRIBUTIONS

Omprakash Karada: Conceptualization, Methodology, Software, Formal Analysis, Investigation, Writing Original Draft.

Prashant Panse: Supervision, Validation, Resources, Writing Review and Editing.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

ETHICS STATEMENT

This study used a publicly available, fully de-identified imaging corpus and involved no human participants or animal subjects. Ethical approval was therefore not required.

DATA AVAILABILITY STATEMENT

The CT–MRI corpus analysed in this study is publicly available through Kaggle. The trained models, federated round logs, SHA-256 audit chain, and figure-generation code are available from the corresponding author on reasonable request.

REFERENCES

[1] Al-Saleh, A., Tejani, G.G., Mishra, S., et al., 2025. A federated learning-based privacy-preserving image processing framework for brain tumor detection from CT scans. *Scientific Reports* 15, 23578. https://doi.org/10.1038/s41598-025-07807-8

[2] Subba, A.B., Sunaniya, A.K., Mukherjee, A., 2025. Multiclass brain tumor detection with attention-embedded CNN framework: advancing toward decentralized deep learning-based health monitoring. *IEEE Journal of Biomedical and Health Informatics*. https://doi.org/10.1109/JBHI.2025.3638154

[3] Anish, J.J., Ajitha, D., 2025. Exploring the state-of-the-art algorithms for brain tumor classification using MRI data. *IEEE Access* 13, 118033–118054. https://doi.org/10.1109/ACCESS.2025.3579727

[4] Lamba, K., Rani, S., 2025. AI-driven emerging trends, challenges, and future directions for brain tumor detection in healthcare. *IEEE Communications Standards Magazine*. https://doi.org/10.1109/MCOMSTD.2025.3645902

[5] Aamir, M., Rahman, Z., Choudhry, N., Bhutto, J.A., Abro, W.A., Zhu, Z., 2025. From CNNs to transformers: a review of evolving deep learning architectures for brain tumor classification. *IEEE Access* 13, 184918–184936. https://doi.org/10.1109/ACCESS.2025.3625607

[6] Yan, Y., et al., 2024. Cross-modal vertical federated learning for MRI reconstruction. *IEEE Journal of Biomedical and Health Informatics* 28(11), 6384–6394. https://doi.org/10.1109/JBHI.2024.3360720

[7] Qazi, E.-u.-H., AL-Ghanem, W.K., Faheem, M.H., Ullah, H., 2026. Federated learning framework for privacy-preserving explainable AI-driven clinical decision-making. *IEEE Journal of Biomedical and Health Informatics*. https://doi.org/10.1109/JBHI.2026.3679499

[8] Gupta, S., Gupta, M., Kumar, R., Abraham, A., 2026. A federated learning and explainable AI framework for privacy-preserving brain tumor diagnosis using multi-institutional MRI data. *IEEE Access* 14, 22630–22643. https://doi.org/10.1109/ACCESS.2026.3660305

[9] Rangaswamyshetty, K.C., Somashankaregowda, M.B., 2026. A techno-analytical insight on federated learning methodologies towards diagnosis of brain tumor. *Bulletin of Electrical Engineering and Informatics* 15(2), 1211–1220.

[10] Appasami, G., Savarimuthu, N., 2025. Federated learning for secure medical MRI brain tumor image classification. *The European Physical Journal Special Topics* 234(15), 4813–4827.

[11] Nath, K.K., Kundu, S., Mohanty, S.N., Mohanty, S., 2026. Federated learning for brain tumor classification: multi-model feature fusion with attention-enhanced CNN. In: *2026 IEEE International Conference on Emerging Computing and Intelligent Technologies (ICoECIT)*, pp. 1–8. IEEE.

[12] Sen, A., Heng, S.H., Tan, S.C., 2026. Privacy-preserving federated learning for MRI-based brain tumour detection using Xception CNN. In: *2026 International Conference on Advances in Artificial Intelligence and Machine Learning (AAIML)*, pp. 124–129. IEEE.

[13] Parvathala, B.R., Malleswari, T.N., Kumari, B.S., 2025. A scalable federated learning framework for efficient brain tumor segmentation with advanced data privacy. In: *2025 5th International Conference on Advancement in Electronics and Communication Engineering (AECE)*, pp. 1054–1059. IEEE.

[14] Salam, M.A., Atef, O., Salam, N.A., Aldawsari, M., 2026. Brain disease diagnosis using federated deep learning. *PeerJ Computer Science* 12, e3545.

[15] Lakshmi Vasanthi, K., Sree Darshne, J., Venkatasubbu, P., Ramasubramanian, P., 2026. A federated multimodal deep learning framework for brain tumor classification using MRI. *Frontiers in Artificial Intelligence* 9, 1754000.

[16] Maharana, R.R., Sahoo, R.R., Dash, D.K., Kumari, T.M., 2026. Decentralized brain tumor detection using transfer learning in federated frameworks. In: *Computing, Communication and Intelligence*, pp. 67–72. CRC Press.

[17] Anoosha, S., Seetharamulu, B., 2025. Federated and deep learning techniques in medical imaging: state-of-the-art innovative approaches for brain tumor segmentation. In: *International Conference on Smart Trends for Information Technology and Computer Communications*, pp. 375–386. Springer Nature Singapore.

[18] Doulani, K., Sharma, S., 2025. Federated learning-enhanced brain tumor prediction on edge networks using MRI imaging. In: *Doctoral Symposium on Human Centered Computing*, pp. 140–151. Springer Nature Singapore.

[19] Bendazzoli, S., Moreno, R., 2025. BraTS-FL: enhancing generalization in brain tumor segmentation via federated learning. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 481–488. Springer Nature Switzerland.

[20] Shaheema, S.B., Jose Kumar, S.J., Selvi, S., Supriya, M., Sajan, R.I., Arul, A.A., 2026. A federated learning based automatic brain tumor segmentation and feature extraction in MRI images. In: *2026 World Conference on Computational Science and Technology (WcCST)*, pp. 53–58. IEEE.

[21] Kumar, M.V., Vedantham, R., 2026. Hybrid deep learning for enhanced brain tumor detection and classification in MRI images. In: *2026 International Conference on Emerging Systems and Intelligent Computing (ESIC)*, pp. 455–461. IEEE.

[22] Rajagopal, S., Shieh, C.S., Shankar, S.S., Maithili, K., Chakrabarti, P., 2026. Privacy-preserving multi-class brain tumor classification using federated learning and deep capsule networks. *Soft Computing*, 1–9.

[23] Chaudhary, N., Thacker, C., 2026. FedBrain-3DMRI: federated self-supervised learning for 3D brain tumor segmentation using SCAFFOLD algorithm. *Journal of Electronics, Electromedical Engineering, and Medical Informatics* 8(2), 672–688.

[24] Chowdhary, C.L., 2026. Federated learning approaches for predicting and modelling brain stroke lesion progression: a review. *Archives of Computational Methods in Engineering*, 1–5.

[25] Hemalatha, D., Chowdary, J.C., Prasad, S., Vijay, V., Rajathi, K., Vijayalakshmi, V., 2026. Federated hybrid CNN-ViT model for brain tumor segmentation. In: *Intelligent and Sustainable Systems*, pp. 234–239. CRC Press.

[26] Chandrasekaran, S., Nethravathi, B., Cholla, R.R., Babu, G.N., Mahesh, T.R., 2026. Enhanced classification of brain MRI images leveraging edge AI strategies and multi-model ensembles. *Neural Computing and Applications* 38(4), 62.

[27] Firdaus, N., Raza, Z., 2026. DenseFedCNN: a federated learning-based framework with client-specific hyperparameter tuning for multi-class brain MRI classification. *International Journal of Machine Learning and Cybernetics* 17(3), 93.

[28] Ghanta, S., Pradhan, A.K., Boyapati, P., Paul, S.P., Rachakonda, L., 2026. Privacy-preserving federated transfer learning for brain tumor multi-classification in AI-driven healthcare. In: *Cybernetic Shield*, pp. 103–118. CRC Press.

[29] Goyal, A.K., Pandey, M., Singh, D.P., Choudhary, J., Haripriya, R., 2024. Collaborative and privacy-enhanced medical image diagnosis using transfer federated learning. In: *International Conference on Innovations in Data Science*, pp. 195–204. Springer Nature Singapore.

[30] Vijayendher, G., Karthikeya, K.S., Madhav, E.J., Kiranmai, T.S., 2025. Enhancing MRI tumor detection: a survey on image upscaling, GAN based data augmentation and federated learning in convolutional neural networks. In: *International Conference on ICT for Sustainable Development*, pp. 321–333. Springer Nature Switzerland.

[31] Markodimitrakis, E., Trivizakis, E., Ioannidis, G.S., Koutoulakis, E., Tsiknakis, M., Papanikolaou, N., Klontzas, M.E., Yaqub, M., Marias, K., 2026. Federated learning on magnetic resonance imaging: a critical review. *Artificial Intelligence Review*.

[32] Shahzad, I., Ouyang, J., Khan, S.U., 2026. FedVC-ADDiM: a federated learning framework for diagnosis of Alzheimer disease using deep learning. *Multimedia Systems* 32(3), 161.

[33] Khan, U.A., Badri, S., Hasan, S.H., 2026. Neuroimaging and machine learning fusion for improved brain tumor diagnosis and prognosis. *Scientific Reports*.