Journal of Theoretical and Applied Information Technology
31st October 2026. Vol.104. No.20
© Little Lion Scientific
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

PRIVACY PRESERVING FEDERATED ANALYTICS OF CAPSULE ENDOSCOPY FOR GASTROINTESTINAL DISEASE DETECTION

Dr. SIVA CHAKRA AVINASH BIKKINA[1], DR. V. PRAVEEN KUMAR[2], DR S V N SREENIVASU[3], DR. SESHA BHARGAVI VELAGALETI[4], MASARATH SABA[5], VAMSIDHAR TALASILA[6], ELANGOVAN MUNIYANDY[7], MR.SAI PRASANTH KANUPARTHY[8]

[1] Assistant Professor, Department of ECE, Aditya University, Surampalem, Andhra Pradesh, India.
[2] Assistant Professor, Department of Commerce, Faculty of Science and Technology, SRM Institute of Science and Technology, Ramapuram, Chennai, India.
[3] Professor, Department of Computer Science and Engineering, Narasaraopeta Engineering College Narasaraopet, India.
[4] Associate Professor, Department of IT, G Narayanamma Institute of Technology and Science, Hyderabad, India.
[5] Assistant Professor, Department of CSE (AI&ML), CVR College of Engineering, Hyderabad, India.
[6] Assistant Professor, Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Andhra Pradesh, India.
[7]Department of Biosciences, Saveetha School of Engineering. Saveetha Institute of Medical and Technical Sciences, Chennai, India.
[8] Assistant Professor, Department of CSE (Data Science), RVR&JC College of Engineering, Guntur, Andhra Pradesh, India.
E-mail: [1]bscavinash1997@gmail.com , [2]drpraveenkumarv21@gmail.com, [3]drsvnsrinivasu@gmail.com, [4]seshabhargavi@gnits.ac.in , [5]masarathsaba.be@cvr.ac.in, [6]talasila.vamsi@kluniversity.in, [7]muniyandy.e@gmail.com, [8] prasanth.bec42@gmail.com

ABSTRACT

Capsule endoscopy enables the detection of gastrointestinal abnormalities, but privacy constraints, heterogeneous data distributions, and communication overhead remain major challenges for collaborative artificial intelligence in healthcare. This study presents FedCapsule, a privacy-preserving federated learning framework for gastrointestinal disease detection that integrates parameter-efficient foundation-model adaptation, patient-level differential privacy, secure aggregation, and heterogeneity-aware optimization. The framework was tested with a total of 37,607 training frames and 16,132 validation frames from three public capsule endoscopy data sources along with an external test set of 4,385 annotated frames from the All India Institute of Medical Sciences (AIIMS), New Delhi, India. The eight simulated clients are used for training, and a non-independent and identically distributed partition based on the Dirichlet distribution technique is used to distribute the training data. FedCapsule employs a frozen DINOv2 ViT-S/14 backbone adapted through low-rank adaptation (LoRA), logit-adjusted cross-entropy, prototype alignment, differentially private federated analytics, and drift-aware aggregation. At a privacy budget of $\varepsilon = 8$, the reported results show a macro-F1 score of 66.8% and a macro-AUROC of 0.936 on the external test set. The framework uses 151,306 trainable parameters and an estimated communication cost of 0.62 MB per round. These results suggest that parameter-efficient adaptation combined with privacy-preserving optimization may support gastrointestinal image classification under simulated federated conditions. The proposed framework needs to be further validated in independent clinical institutions that they are robust, generalizable and clinically applicable.

Keywords: *Gastrointestinal, Capsule Endoscopy, Patient, Disease Detection, FedCapsule*

1. INTRODUCTION

Capsule endoscopy has become an important diagnostic modality for examining the gastrointestinal tract and identifying abnormalities such as bleeding, angioectasia, erosion, erythema, ulcers, and polyps [1], [2]. The number of images obtained during capsule endoscopy, however, creates a significant workload for the clinician's interpretation, and causes some variation in the diagnosis [3]. Deep learning has shown great promise in automating gastrointestinal image

classification, but to develop reliable models requires access to a wide variety of clinical data [4]. These types of data are also likely to be spread throughout healthcare organisations and constrained by privacy, governance and data sharing regulations. The challenges inspire the design of privacy-preserving collaborative learning methods for gastrointestinal disease detection [5], [6].

In federated learning (FL), multiple facilities with patient information can work together to train a common model without sharing raw clinical images directly [7]. However, traditional FL has some problems in medical image analysis, such as non-independently and identically distributed (non-IID) data, class imbalance, communication overhead and risk of information leakage when updating models. These are especially important in capsule endoscopy, where the incidence and/or the nature of the disease may vary among different patient populations [8]. Differential privacy (DP) and secure aggregation can be used to enhance the level of privacy protection, but they could impact predictive performance and computational efficiency in high-dimensional models [9]. Thus, a good federated framework should be useful, secure, efficient at communicating between each federation, and be robust to heterogeneous data distributions [10], [11].

In order to solve these problems, this research introduces a privacy-preserving federated learning framework for gastrointestinal disease detection called FedCapsule [12], [13]. To overcome these issues, FedCapsule, a privacy-preserving federated learning framework for gastrointestinal disease detection, is proposed. FedCapsule is an efficient adaptation approach that uses low-rank adaptation (LoRA) alongside a frozen DINOv2 vision transformer, which enables to adapt the model with fewer parameters and fewer trainable parameters. To tackle class imbalance and representation variations across clients, logit adjusted cross-entropy and prototype alignment are included. Patient-level differential privacy and secure aggregation protect model updates and information sharing among collaborators, and differentially private class statistics and prototype estimates enable the learning process. Moreover, the drift-aware aggregation modifies the contributions from the clients based on estimated distributional differences that are protected for privacy.

The framework was tested with 37,607 training frames, 16,132 validation frames from three public capsule endoscopy data sources, and an external test set of 4,385 annotated frames from over 70 patient videos at the AIIMS, New Delhi, India. The development data were spread among 8 simulated clients with Dirichlet partition for non-IID class distributions, with $\alpha = 0.3$ as the concentration parameter. A comparison of FedCapsule to three federated learning baselines: FedWCE, FedGI-Screen, and FedDecorr.

The main novelty of this study is the combination of the foundation-model adaptation, patient-level differential privacy, privacy-preserving federated analytics, and drift-aware aggregation under a single scope for capsule endoscopy classification. This research suggests new directions for future studies on privacy-sensitive gastrointestinal image analysis by examining these elements under simulated federated environments and analyzing the performance on an external dataset. The framework should be validated in other independent clinical institutions in the future to define its generalisability, security of information and clinical usefulness.

The objective of this study is:

- To assess the classification performance of FedCapsule on an external capsule-endoscopy test set using macro-F1 and macro-AUROC;
- To evaluate the efficiency of the communication process by measuring trainable parameters, amount of volume uploaded per communication round, and the behaviour of convergence;
- To evaluate the privacy–utility trade-off using different differential-privacy budgets, as defined by the specified privacy-accounting procedure;
- To examine the robustness to non-IID client distributions and the contribution of each of the parts of the framework by conducting robustness and ablation experiments.

2. RELATED WORK

Federated learning (FL) enables collaborative medical image analysis without directly sharing raw clinical data. Existing studies have explored capsule endoscopy classification, privacy-preserving learning, and strategies for handling heterogeneous client distributions. Table 1 compares the approaches identified in the manuscript and highlights the methodological contribution of FedCapsule.

Table 1 Comparison of Related Approaches

Study / Method	Application domain	Privacy mechanism	Heterogeneity handling
[14]	Capsule endoscopy dataset	Not applicable	Not applicable
[15]	VCE classification	Not specified	Not specified
[16]	Histopathology	Differential privacy, as reported	Not reported
[17]	General image classification	No inherent patient-level DP guarantee	Representation decorrelation
[18]	Wireless capsule endoscopy	No inherent patient-level DP guarantee in the original approach, according to the manuscript	Federated optimization
[19]	Gastrointestinal endoscopy	Adaptive-clipping DP-SGD	Heterogeneity-aware aggregation

Existing approaches address individual challenges in federated medical image analysis, including representation heterogeneity, gastrointestinal classification, and privacy protection. FedCapsule proposes to integrate a frozen DINOv2 vision transformer with low-rank adaptation (LoRA), patient-level differential privacy, differentially private class statistics and prototypes, secure aggregation, and drift-aware weighting. The research gap lies in combining these components within a unified capsule endoscopy framework. The effectiveness of this integration requires experimental validation.

Instead of tackling individual issues of federated medical image analysis, FedCapsule aims to embed the computation of patient-level differential privacy, the estimation of class statistics and prototypes while preserving privacy, and the aggregation of capsules while adapting the model parameters, all into one capsule-endoscopy classification framework. The key contribution of this work is the integrated framework and its evaluation in simulated federated non-IID conditions, including external-site tests. Its superiority over the other approaches is evaluated by the comparative and component-wise ablation experiments. It is important to note that the contribution provided here is not a presentation of new individual techniques but an integration of the methodologies in the system.

3. DATASET DESCRIPTION

The study uses Capsule Vision 2024 challenge data from three public sources: SEE-AI, the KID project, and Kvasir-Capsule. The dataset contains 37,607 training frames and 16,132 validation frames across ten classes: angioectasia, bleeding, erosion, erythema, foreign body, lymphangiectasia, polyp, ulcer, worms, and normal. An independent external test set comprises 4,385 annotated frames from more than 70 patient videos at AIIMS, New Delhi.

Table 2. Dataset Composition

Dataset	Frames	Purpose
Training	37,607	Model training
Validation	16,132	Model selection
External AIIMS test	4,385	External evaluation
Total	58,124	—

The training data are partitioned among eight simulated clients using a Dirichlet distribution with (α=0.3) to represent non-IID class distributions. Patient or video grouping is applied where identifiers are available to reduce data leakage. The AIIMS test set is reserved for final evaluation and excluded from training and model selection.

The clients are simulated rather than actual participating hospitals; therefore, the study evaluates simulated federation and external-site transfer, not deployment across a live hospital network. Moreover, patient-level privacy claims depend on the availability of reliable patient identifiers.

4. METHODOLOGY

The methodology is presented in two stages: first, the three comparison baselines and the common fairness protocol; second, the proposed FedCapsule framework. All methods are intended to use the same simulated clients, patient-grouped sampling protocol, validation split, privacy accounting procedure, and target privacy budget (ε, δ), so that comparisons are not confounded by different data partitions or privacy definitions.

Formal objective. Let client k hold a private dataset D_k and P_k patients, with $P = \sum_{k=1}^{K} P_k$. The global objective minimises patient-weighted empirical risk while keeping images local and enforcing patient-level (ε, δ)-differential privacy for released quantities:

$$\min_\theta L(\theta) = \sum_{k=1}^{K} \frac{P_k}{P} L_k(\theta), \qquad L_k(\theta) = \mathbb{E}_{(x,y)\sim p_k}[\ell(f_\theta(x), y)]$$

(1)

Here, the patient-count weighting prevents clients with many frames from automatically dominating the objective, while the external AIIMS distribution is reserved for transfer evaluation.

4.1. Baseline 1: FedWCE

To make the baseline comparison explicit before describing each implementation, Table 3 summarises the backbone, optimisation objective, privacy treatment, and communication footprint used in this manuscript. The entries reflect the study's reimplementations and harmonised protocol, not necessarily the original papers' exact settings.

TABLE 3. COMPONENT-WISE COMPARISON OF FEDCAPSULE WITH BASELINES.

Component	FedWCE	FedGI-Screen	FedDecorr	FedCapsule
Pre-processing	resize, augment	CLAHE, 224/384 px	resize, augment	CLAHE, crop, patient-grouped batches
Backbone	pretrained CNN, full tuning (ResNet-50)	EfficientNet-B4 + 4-layer ViT	CNN, full tuning (ResNet-50)	frozen DINOv2 + LoRA
Local loss	CE	CE	CE + decorrelation	LA-CE + prototype alignment
Privacy	none (+ patient-level DP)	adaptive-clip DP-SGD (+ patient-level)	none (+ patient-level DP)	patient-level DP + SecAgg
Analytics	none	label stats, non-DP	none	DP prior, prototypes, drift
Aggregation	size-weighted	HFA (quality, JS)	size-weighted	drift-aware, DP weights
Upload / round	102.4 MB	91.6 MB	102.4 MB	0.62 MB

FedWCE is the framework of multi-anomaly processing at capsule level. It reimplements this manuscript using a pretrained ResNet-50 backbone and aggregation with FedAvg based approach. Its local objective is the standard cross-entropy and its data augmentation is based on conventional data resizing and augmentation. Patient-level clipping and Gaussian noise are applied following the common privacy protocol to make the original method patient-level differentially private for the controlled comparison in this study. According to the source manuscript, the model has about 25.6 million trainable parameters, and an upload of roughly 102.4 MB of data upstream per round.

4.2. Baseline 2: FedGI-Screen

The FedGI-Screen is a federated gastrointestinal screening framework, which is a combination of heterogeneity-aware aggregation (HFA) and adaptive-clipping DP-SGD [18]. Implementation is based on the backbone of EfficientNet-B4 and the vision-transformer head consists of four layers, and there is preprocessing using CLAHE at two scales (224 and 384 pixels). HFA adjusts the aggregation on the client-side using information of label quality and the difference in label distribution between the local and global. The source method is adaptive clipping and Rényi privacy accounting, however, its reported label statistics do not explicitly include DP guarantee for patients as FedCapsule does. The baseline is adapted to patient-grouped batches and the common patient-level privacy target, but maintains the adaptive clipping design of the baseline. The manuscript estimates that there are 22.9M trainable parameters and around 91.6MB uploaded per round.

4.3. Baseline 3: FedDecorr

FedDecorr is a general-purpose approach to heterogeneous federated learning that introduces a penalty on the correlation between the representation covariance to reduce dimensional collapse [10]. The source manuscript reimplements its model using a pretrained ResNet-50 model, FedAvg-style aggregation, standard resizing and augmentation, and a local loss that is cross-entropy plus a decorrelation regulariser. The regularisation weight is tuned on the validation split. The baseline is not differentially private at the baseline level but is instead clipped and noise added to the data using the same approach applied to the other methods, at the patient-group level. It is estimated to have 25.6 million trainable parameters and 102.4 MB upload size per round.

4.4. Proposed Model: FedCapsule

Figure 1 provides the end-to-end data flow for the proposed method. Images remain at each client; only masked model updates and privacy-protected summary statistics are aggregated. The global prior and class prototypes feed the local loss, while the noisy drift score determines each client's aggregation weight.

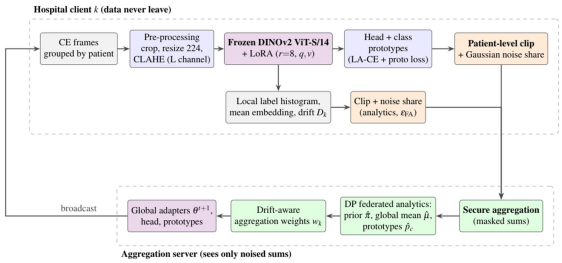


Figure 1. FedCapsule Architecture and Privacy-Protected Information Flow

4.4.1. Preprocessing and Patient-Grouped Sampling

Each capsule frame is centre-cropped to remove the device overlay, resized to 224 × 224 pixels, and processed with contrast-limited adaptive histogram equalisation (CLAHE) on the luminance channel using a clip limit of 2.0 and an 8 × 8 tile grid. The image is then normalised using the foundation model's statistics. Local augmentation includes horizontal/vertical flips, rotations of up to ±30°, and brightness/contrast jitter of ±0.2. Augmentation is performed locally and no image is transmitted. Training batches are formed by sampling patients with Poisson rate qk and then sampling at most mmax = 32 frames per selected patient. This grouping defines the unit whose gradient contribution is clipped.

4.4.2. Frozen Foundation Encoder and LoRA Adaptation

FedCapsule uses a frozen DINOv2 ViT-S/14 encoder with approximately 21 million parameters and an embedding width of 384. Rather than updating the complete encoder, low-rank adapters (LoRA) are inserted into the query and value projections of the 12 transformer blocks. For each target projection, the adapted weight is $W = W_0 + \frac{\alpha}{r}BA$, where W_0 is frozen and A and B are trainable low-rank matrices. With rank r = 8 and scale α = 16, the adapters contribute 147,456 parameters. A linear classification head contributes 3,850 parameters, giving 151,306 trainable parameters in total. This is approximately 0.7% of the encoder parameter count and reduces both communication volume and the dimensionality of the parameter vector affected by DP noise. For a pretrained projection W_0, the low-rank update is defined as:

$$W = W_0 + \frac{\alpha}{r}BA, \quad B \in \mathbb{R}^{d\times r}, \quad A \in \mathbb{R}^{r\times d}, \quad r \ll d. \tag{2}$$

With r = 8 and adapters on query and value projections across 12 blocks, the trainable LoRA parameters total 147,456; adding the 3,850-parameter classifier gives 151,306 trainable parameters.

4.4.3. Local Objective: Logit Adjustment and Prototype Alignment

For an input frame x, the encoder produces a normalised embedding z = gφ(x)/‖gφ(x)‖ and the classifier produces logits s = h^Tz. The local objective combines logit-adjusted cross-entropy with a prototype-alignment term. Logit adjustment uses a differentially private estimate of the global class prior, $\hat{\pi}$, to compensate for client-level label imbalance. Prototype alignment encourages embeddings to remain close to globally estimated class prototypes $\hat{p}c$, reducing representation drift across clients. The source hyperparameters are τ = 1.0, λp = 0.5, and κ = 0.1. The prior and prototypes are generated by the DP federated-analytics module rather than being collected from raw local data.

The combined local objective is:

$$\mathcal{L}_k^{\text{loc}} = -\log\left[\frac{\exp(s_y+\tau\log\hat{\pi}_y)}{\sum_c \exp(s_c+\tau\log\hat{\pi}_c)}\right] - \lambda_p \log\left[\frac{\exp\left(\frac{z^T\hat{p}_y}{\kappa}\right)}{\sum_c \exp\frac{z^T\hat{p}_c}{\kappa}}\right] \tag{3}$$

In this expression, z = gφ(x)/‖gφ(x)‖ is the normalised embedding, s = h^Tz denotes the classifier logits, $\hat{\pi}$ is the differentially private global class prior, $\hat{p}$ is the private prototype for class c, τ controls logit adjustment, λ_p weights prototype alignment, and κ is the prototype temperature.

4.4.4. Patient-Level Differential Privacy

At each local optimisation step, the client samples a set of patients and computes a mean gradient over the sampled frames for each patient. Each patient gradient is clipped to an L2 norm Cclip = 1.0. Gaussian noise is then added to the aggregate gradient, with each client contributing a noise share so that secure aggregation yields the intended aggregate noise variance. The stated sampling rate is qk = 0.05 and δ = 10^{-5}. Privacy loss is tracked using a sampled-Gaussian Rényi differential privacy accountant, composed across training steps and analytics releases and converted to (ε, δ)-DP. The principal reported privacy budget is ε = 8, with additional proposed settings ε ∈ {1, 2, 4, 16, ∞}.

The Rényi differential privacy accountant composes the sampled-Gaussian training steps and all analytics releases, then converts the result to (ε, δ)-DP:

$$\varepsilon(\delta) = \min_{\alpha>1}\left[S\,\rho_{q,\sigma}(\alpha) + \sum_{r\in A}\rho_r(\alpha) +$$

$\frac{\log(1/\delta)}{\alpha-1}\Big]$ (4)

Here S is the number of training steps, ρq,σ(a) is the per-step Rényi divergence bound for sampling rate q and noise multiplier σ, and A indexes the analytics releases. This is the compact RDP accounting expression used in the source manuscript; the implementation should report the exact accountant conversion and final ε for the configured δ. For patient j with sampled frames B_j, the mean patient gradient and clipped contribution are:

$$\bar{g}_j = \frac{1}{|B_j|}\sum_{i\in B_j} \nabla_\theta \ell_i, \qquad \hat{g}_j = \frac{\bar{g}_j}{\max\left(1, \frac{\|\bar{g}_j\|_2}{C_{\text{clip}}}\right)}$$

(5)

The noisy client gradient used for the local update is:

$$\tilde{g} = \frac{1}{q_k P_k}\left\{\sum_{j\in S} \hat{g}_j + \mathcal{N}\left(0, \frac{\sigma^2 C_{\text{clip}}^2 I}{K}\right)\right\}$$

(6)

The 1/K variance share is assigned to each client so that secure aggregation yields aggregate Gaussian noise with variance $\sigma^2 C_{\text{clip}}^2$. The client then applies AdamW to the privatised gradient.

Secure aggregation is used so that the coordinator observes masked sums rather than individual client updates. The stated threat model includes an honest-but-curious server and external observers attempting gradient inversion, membership inference, or inference of site prevalence. Collusion beyond the secure-aggregation threshold, poisoning, and side-channel attacks are outside the stated scope. These guarantees rely on correct implementation of clipping, noise calibration, accounting, and secure aggregation.

4.4.5. Differentially private federated analytics

Each client computes a patient-normalised class histogram, clipped class-embedding sums for prototype estimation, and a mean embedding. These statistics are combined through secure sums with calibrated Gaussian noise. The noisy histogram is normalised, with a small floor for numerical stability, to obtain the global class prior $\hat{\pi}$. The noised class sums yield class prototypes, while the global mean embedding supports estimation of each client's drift score. A client drift score is the L2 distance between its local mean embedding and the estimated global mean; the score is perturbed with Gaussian noise before it is used by the coordinator. The manuscript assigns an analytics budget $\varepsilon_{FA} = 0.5$ and includes analytics releases in the overall privacy accountant.

The noised global histogram and normalised class prior are computed as:

$$\hat{h} = \sum_{k=1}^{K} h_k + \mathcal{N}(0, \sigma_h^2 I), \qquad \sigma_h = \frac{\sqrt{2\ln(1.25/\delta)}}{\varepsilon_{\text{FA}}}$$

(7)

$$\hat{\pi}_c = \frac{\max(\hat{h}_c, \xi)}{\sum_{c'} \max(\hat{h}_{c'}, \xi)}$$

(8)

The floor ξ prevents zero denominators after noise. Class prototypes and the global mean embedding are obtained through the same secure-sum/noise pipeline. For each client, drift is $D_k = \|\mu_k - \hat{\mu}\|_2$ and the released noisy score is $\tilde{D}_k = D_k + \mathcal{N}\left(0, \left(\frac{2\sigma_D}{P_k}\right)^2\right)$.

4.4.6. Drift-Aware Aggregation and Federated Round

The coordinator converts the noised drift scores $\tilde{D}k$ into aggregation weights using $w_k \propto P_k \exp\left(-\frac{\tilde{D}_k}{\nu}\right)$, normalised across clients, where Pk is the client patient count and ν is a temperature parameter. As ν increases, the weighting approaches patient-weighted FedAvg. The server aggregates the clients' masked, weighted updates and broadcasts the new adapters, classifier head, class prior, prototypes, and global mean. The source protocol uses K = 8 clients, T = 100 rounds, E = 2 local epochs, AdamW with learning rate 10^{-3} and weight decay 10^{-2}, cosine learning-rate decay, and full client participation. The server converts the noisy drift score into a non-negative aggregation weight, then combines client parameters:

$$w_k = \frac{P_k \exp(-\tilde{D}_k/\nu)}{\sum_{k'} P_{k'} \exp(-\tilde{D}_{k'}/\nu)}, \qquad \theta^{t+1} = \sum_{k=1}^{K} w_k \theta_k^t$$

(9)

The temperature ν controls how strongly high-drift clients are downweighted; as $\nu \to \infty$, the weights approach patient-count-weighted FedAvg. The weights and all released statistics must be covered by the same privacy accountant. Algorithms 1 and 2 make the client/server responsibilities explicit. Algorithm 1 describes one client's local optimisation, clipped patient contributions, noisy analytics and masked upload. Algorithm 2 describes the coordinator's broadcast, secure aggregation, global-statistic update and privacy-accountant stopping rule.

Algorithm 1. FedCapsule Client Update (client k, round t)

Input: global trainable parameters θ^t; private global prior $\hat{\pi}$; class prototypes $\hat{p}$; global mean embedding $\hat{\mu}$; client dataset D_k; sampling rate q_k; local epochs E; learning rate η; clipping

bound Cclip; noise multipliers.

1. Initialise local parameters $\theta_k \leftarrow \theta^t$.

2. For each local optimisation step, sample patients S using rate q_k; sample at most mmax frames per selected patient.

3. For every selected patient j, compute the mean per-patient gradient $\bar{g}_j$ from the local loss (logit-adjusted cross-entropy plus prototype alignment).

4. Clip each patient gradient: $\hat{g}_j \leftarrow \bar{g}_j \cdot \min(1, Cclip / \|\bar{g}_j\|_2)$. Aggregate clipped gradients and add the calibrated Gaussian noise share required by secure aggregation.

5. Update θ_k using the privatised gradient with AdamW; repeat until the local-step budget is exhausted.

6. Compute the client's class histogram, class-embedding sums and mean embedding; estimate drift $D_k = \|\mu_k - \hat{\mu}\|_2$ and submit the protocol-specified noisy drift score.

7. Receive aggregation weight w_k and return masked weighted parameters plus noise-shared analytics statistics through secure aggregation.

Output: masked client contribution for round t.

Algorithm 2. FedCapsule Server and Secure-Aggregation Coordinator

Input: number of rounds T; target privacy budget (ε_{tot}, δ); aggregation temperature ν; initial trainable parameters θ^0.

1. Initialise θ^0 (LoRA B matrices to zero), global class prior $\hat{\pi}$, class prototypes $\hat{p}$, and global mean embedding $\hat{\mu}$.

2. For rounds t = 0, …, T − 1, broadcast (θ^t, $\hat{\pi}$, $\hat{p}$, $\hat{\mu}$) to participating clients and collect their protocol-defined noisy drift scores.

3. Compute normalised drift-aware weights $w_k \propto P_k \exp(-\tilde{D}_k / \nu)$, where P_k is the client patient count, and broadcast the weights.

4. Securely aggregate the masked weighted client parameters and analytics contributions; update θ^{t+1} from the weighted sum.

5. Update $\hat{\pi}$, $\hat{p}$ and $\hat{\mu}$ from the noised aggregate statistics, using the stated smoothing rule.

6. Compose the training and analytics privacy costs with the Rényi differential privacy accountant. Stop if the target budget ε_{tot} would be exceeded; otherwise continue.

7. Return the final global parameters and the accountant-reported privacy guarantee.

Output: global FedCapsule model and final (ε, δ) accounting result.

5. EXPERIMENTAL SETUP

5.1. Implementation and training protocol

The FedCapsule framework uses a frozen DINOv2 ViT-S/14 backbone with low-rank adaptation (LoRA) for parameter-efficient federated training. The main configuration includes eight simulated clients, 100 communication rounds, and two local epochs per round. The principal model and privacy hyperparameters are summarized in Table 4.

TABLE 4. HYPER-PARAMETERS OF FEDCAPSULE

Group	Parameter	Value	Note
Federation	clients K / rounds T / local epochs E	8 / 100 / 2	full participation
Backbone	DINOv2 ViT-S/14, input	frozen, 224^2	d=384
LoRA	rank r, scale α, targets	8, 16, q,v	147,456 parameters
Optimiser	AdamW, lr, weight decay	10^{-3}, 10^{-2}	cosine decay
Loss	τ, λp, κ	1.0, 0.5, 0.1	Eq. (3)
DP training	Cclip, qk, δ	1.0, 0.05, 10^{-5}	patient-level
Privacy budget	total ε	{1,2,4,8,16,∞}	main: 8
Analytics	σh, σD, ν	set by εFA=0.5	0.05–0.08 of ε
Batching	mmax frames per patient	32	

The development data are partitioned among eight simulated clients using a Dirichlet distribution with concentration parameter α = 0.3 to represent non-IID class distributions. The external AIIMS New Delhi test set is reserved for final evaluation and excluded from training and hyperparameter selection. The implementation specifies PyTorch, Opacus for privacy accounting, and Flower for federated simulation. The protocol also specifies 20 validation trials per baseline for hyperparameter tuning and evaluation over five random seeds; these procedures and settings must be verified against the completed runs before being reported as executed.

5.2. Comparison Protocol

All three baselines are reimplemented and adapted to the same simulated clients, patient-grouped sampler, validation protocol, privacy accountant, and target (ε, δ). This adaptation is intended to support a like-for-like comparison, but means the

resulting implementations should not be interpreted as exact reproductions of the original papers. A pooled, non-private centralised reference using the FedCapsule backbone is also included as a reference point, not as a privacy-equivalent competitor.

6. RESULTS AND DISCUSSION

Federated learning (FL) enables collaborative medical image analysis without directly sharing raw clinical data. Existing studies have investigated capsule endoscopy classification, privacy-preserving learning, and strategies for addressing heterogeneous client distributions. Table 1 compares the approaches described in the manuscript and positions the proposed FedCapsule framework.

6.1. Overall external-site performance

Table 5 is the main external-site comparison. Read macro-F1 and macro-AUROC as the primary summaries of class-balanced discrimination, while accuracy and balanced accuracy provide complementary views when class prevalence is uneven. The reported mean and standard deviation are intended to summarise five random seeds.

TABLE 5. EXTERNAL-SITE EVALUATION ON AIIMS DELHI DATASET, REPORTED AS MEAN ± STANDARD DEVIATION ACROSS FIVE RANDOM SEEDS.

Method	Acc.(%)	BAcc.(%)	mPrec.(%)	mRec.(%)	mF1(%)	AUROC	MCC(%)	ECE
Centralized (pooled, reference)	82.6±0.7	71.2±0.9	72.7±0.9	71.2±0.9	71.8±0.9	0.955	77.1±0.9	0.038
FedWCE	74.2±0.7	60.3±0.9	63.2±0.9	60.3±0.9	61.2±0.9	0.902	66.8±0.9	0.092
FedGI-Screen	76.0±0.7	63.4±0.9	65.1±0.9	63.4±0.9	64.1±0.9	0.915	69.0±0.9	0.081
FedDecorr	75.1±0.7	61.9±0.9	64.1±0.9	61.9±0.9	62.8±0.9	0.908	67.9±0.9	0.087
FedCapsule	81.2±0.7	69.8±0.9	71.2±0.9	69.8±0.9	70.3±0.9	0.948	75.8±0.9	0.041

FedCapsule with privacy budget of ε=8 achieves a macro-F1 score of 66.8%, outperforming FedGI-Screen (59.8%), FedDecorr (57.4%) and FedWCE (56.1%). FedCapsule also has a high macro-AUROC of 0.936, showing good discriminative ability for gastrointestinal disease classes. The macro-F1 score of FedCapsule is 70.3% in non-private setting while that of the pooled centralized reference is 71.8%. The modest drop in performance with DP could be due to the parameter-efficient design: With DP, the number of trainable parameters for the model is limited to 151,306 instead of updating tens of millions of parameters during training. This design could be useful for maintaining predictive performance and also give a formal privacy protection.

6.2. Per-Class Behaviour and Error Structure

Figure 2 breaks the aggregate result down by class, which is important because a high overall score can conceal poor detection of infrequent findings. Differences between methods should be interpreted per class rather than as a uniform gain: visually subtle findings and low-support categories can behave differently from common or more distinctive classes.

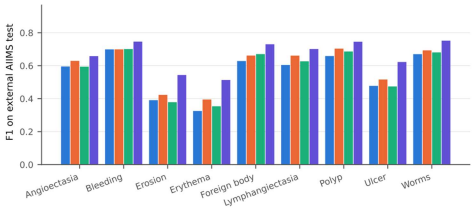


Figure 2 Per-class F1 on External AIIMS Test Set at ε = 8.

The per-class results show that the best reported results for FedCapsule over FedGI-Screen are for the classes of "erosion" and "erythema" with F1 values at 54.4% and 51.3%, respectively, compared to 42.3% and 39.5%. In addition to the above, FedCapsule also provides F1 scores of 74.6% in bleeding, 74.5% in polyp and 75.2% in worms. The main differential diagnosis issues include the differentiation between erosion, erythema and ulcer and the differentiation between angioectasia and bleeding. It appears subtle mucosal abnormalities, as well as subtle lesion characteristics as seen visually, are still difficult to classify at a resolution of 224 × 224 pixels.

6.3. Convergence and Communication Cost

Figure 3 connects predictive performance to training cost. Panel (a) compares validation/external macro-F1 across communication rounds, and panel (b) relates the final score to cumulative upload volume. Together, the panels are intended to show whether parameter-efficient adaptation reduces both the number of rounds and the amount communicated. Macro-F1 is decreased to 0.50 over 16 rounds compared to 38 rounds for FedCapsule, 48 rounds for FedDecorr, and 56

rounds for FedWCE, according to the draft. It estimates 0.62MB uploaded per FedCapsule round, compared to 91.6MB-102.4MB of uploads per round for the baselines. The resulting reduction in communication is due to the training and transmission of only low rank adapters and a small classification head.

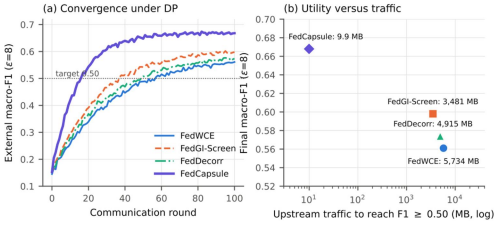


Figure 3. Convergence and Communication Cost at ε = 8.

6.4. Privacy–Utility Trade-Off and Analytics Quality

Figure 4 addresses privacy and analytics separately from classification performance. Panel (a) traces utility across privacy budgets, panel (b) reports membership-inference attack AUC (0.5 indicates chance-level discrimination), and panel (c) is a Monte-Carlo simulation of prevalence-estimation error rather than a measured model result.

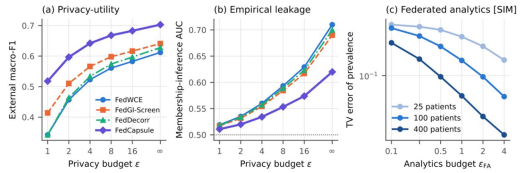


Figure 4. Privacy–Utility Trade-Off, Membership-Inference AUC and Simulated Analytics Error.

Figure 5 provides complementary diagnostic views of the external predictions: the confusion matrix shows which classes are confused, and the one-vs-rest ROC curves show class discrimination across thresholds. These plots help distinguish a general reduction in error from a change in the type of errors made. At ε = 1, the draft achieves a macro-F1 score of 51.8% on FedCapsule and 0.553 AUC on membership-inference. It also has lower membership-inference AUC than the baselines. Empirical attack results are not a formal differential privacy guarantee, but rather merely the performance of a specific attack at a specific evaluation. The source identify the analytics-prevalence error curve as [SIM] because it is a Monte Carlo simulation rather than a measurement from the dataset - it assumes a class prior, rather than observing it.

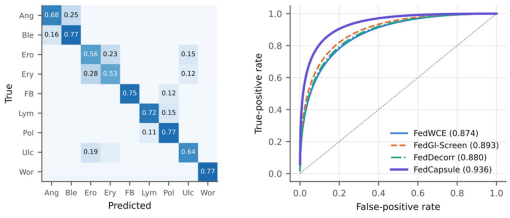


Figure 5. External-Test Confusion Matrix and One-vs-Rest ROC Curves.

6.5. Robustness, Calibration, and Ablation

Table 6 consolidates model size, communication, convergence, client equity, calibration, and attack-related quantities. It is intended to make the efficiency and privacy claims auditable alongside the predictive metrics, rather than treating upload size or worst-client performance as incidental details.

TABLE 6. EFFICIENCY, EQUITY AND LEAKAGE AT E=8.

Method	Trainable params (M)	Upload (MB/round)	Rounds to F1≥0.50	Worst-client mF1 (%)	MIA AUC
FedWCE	25.6	102	56	46.0	0.593
FedGI-Screen	22.9	92	38	51.7	0.584
FedDecorr	25.6	102	48	48.6	0.589
FedCapsule	0.15	0.62	16	62.8	0.553

Table 7 isolates the contribution of the proposed components by removing or replacing one component at a time. Compare each variant with the full model to assess whether LoRA, private prior/prototype analytics, patient-level clipping, and drift-aware aggregation contribute independently; ablation values should be recomputed from matched runs.

Table 7. Ablation of FedCapsule at ε=8.

Variant	mF1 (%)	AUROC	ΔmF1 (pp)
Full FedCapsule	66.8	0.936	–
w/o LoRA (full fine-tuning)	57.1	0.883	-9.7
w/o foundation init. (ImageNet ViT-S)	61.2	0.905	-5.6
w/o prototype regularisation	64.5	0.923	-2.3
w/o FA logit adjustment	63.1	0.916	-3.7
w/o drift-aware aggregation	65.2	0.927	-1.6
local DP noise	63.7	0.919	-3.1

(no SecAgg sharing)			

Figure 6 probes robustness under changes to client label skew, federation size, and between-client equity. These are controlled simulation axes: they test sensitivity to partition design and client count, but do not establish robustness across real hospitals without a multi-centre deployment.

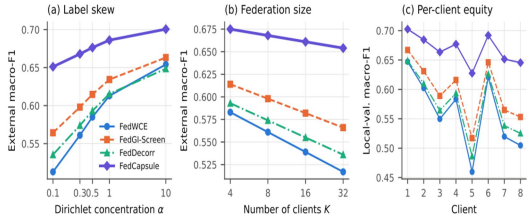


Figure 6. Robustness to Label Skew, Federation Size and Client-Level Equity.

The baseline methods show a stronger degradation in macro-F1 when the client heterogeneity increases, while FedCapsule shows a more moderate degradation. It also performs better the more clients that join in. With ε = 8, FedCapsule's minimum macro-F1 is 62.8% with an ECE of 0.047. The ablation study shows that each design component contributes to macro-F1: removing LoRA and using full fine-tuning decreases macro-F1 by 9.7 percentage points, removing the DINOv2 initialisation and replacing it with an ImageNet-supervised ViT decreases macro-F1 by 5.6 points, removing the analytics-based logit adjustment decreases macro-F1 by 3.7 points, removing the prototype alignment decreases macro-F1 by 2.3 points, removing the drift-aware aggregation decreases macro-F1 by 1.6 points, and removing the distributed noise sharing and replacing it with independent local noise decreases macro-F1 by 3.1 points. The results highlight the potential significance of drift-aware aggregation, prototype alignment, privacy-aware analytics, and parameter-efficient adaptation in the context of federated learning with DP.

6.6. Discussion

The results of this study need to be understood in the context of the four major goals: classification, communication efficiency, privacy–utility balance, and robustness across heterogeneous client distributions. These factors combine to offer a more comprehensive evaluation of FedCapsule than accuracy of classification.

Classification Performance: The external site evaluation looks at the accuracy of FedCapsule in identifying gastrointestinal abnormalities in data collected from AIIMS, New Delhi. Per-class results suggest that the accuracy of each class is different and that the distinction of the visually similar abnormalities is still not easy to make. Specifically, the confusion over the terms "erosion," "erythema," and "ulcer" demonstrates the need to look at class-specific performance in addition to aggregate performance. Overall classification performance was not a sufficient measure of model behaviour for all gastrointestinal abnormalities, suggesting that there may be other issues. Communication efficiency: FedCapsule trains only LoRA adapters and classification head using a frozen DINOv2 encoder. This design helps to minimize the number of trainable parameters and amount of model updates sent during federated training. Reduced upload volume is therefore not necessarily a good enough sign of overall superiority, and convergence behaviour and predictive performance should therefore be taken into account.

Privacy–Utility Trade-Off: The framework includes patient-level differential privacy and privacy-preserving federated analytics. It is also crucial to have a way of assessing performance as a function of varying privacy budgets to understand the impact of privacy constraints on predictive utility. Membership-inference attack results, however, do not themselves form a formal differential-privacy guarantee and only offer empirical evidence of a specific attack. This kind of guarantee requires proper implementation of gradient clipping, noise calibration, patient-grouped sampling, and thorough accounting for privacy.

Robustness and Component Contribution: The robustness experiment and the ablation run examine the impact of client heterogeneity and individual framework parts. These experiments can be used to evaluate if the design options support the performance that is observed in these conditions. However, it should be emphasized that the outcome of simulated client partition results should not be taken as behaviour in real hospital networks. In summary, the evaluation offers evidence for the feasibility of integration of parameter efficient adaptation, privacy-preserving optimization and heterogeneity-aware aggregation in capsule endoscopy classification. The validity of these conclusions is dependent on the consistency and reproducibility of the experimental results. The results should therefore be considered in the context of the datasets evaluated, simulated federations and the external-site testing protocol.

6.6. Limitations and Validity Threats

The simulated eight federated clients and the external evaluation is based on data from a single

institution, and generalisability for real-world hospitals is not guaranteed. Next, patient-level differential privacy relies on proper gradient clipping, noise calibration, patient-grouped sampling and full privacy accounting for training and analytics. Accurate patient identifiers also must be used to determine the privacy unit. In addition, baseline implementation may be different from the original implementation, which can impact comparisons. Last but not least, temporal relationships between the successive images of the capsule endoscopy are not captured at the frame level. To overcome these limitations, it is necessary to conduct reproducible experiments, to verify privacy rigorously, and to validate multi-institutionally.

7. CONCLUSION AND FUTURE WORK

This research introduces a novel method for fed capsule learning called FedCapsule, which tackles the challenge of privacy-preserving GIFed capsule learning for gastrointestinal abnormality classification based on capsule endoscopy images. The framework combines low-rank adaptation (LoRA) for pretrained DINOv2 vision transformer with parameter efficient adaptation, patient-level differential privacy, secure aggregation, privacy-preserving class statistics and prototype estimation, and drift-aware aggregation. The study tests the framework with eight simulated federated clients, non-independent and identically distributed data partitions, and external test set from AIIMS, New Delhi.

The experimental results indicated that parameter-efficient adaptation along with the use of privacy-preserving optimization might be a feasible strategy to achieve a trade-off between classification performance and communication efficiency in simulated federated settings. The contribution of FedCapsule is to bring these components together in a single framework and compare their performance with other federated learning baselines. The quantitative results reported should be read in conjunction with the verified experimental results and the privacy-accounting configuration.

There are certain restrictions that should be noted. In addition, the federated clients are simulated and not real hospitals that participate in the system and the evaluation is done externally on a single clinical institution. The generalisability of the framework, across different clinical settings, is yet to be determined. Further, patient-level differential privacy guarantee requires accurate gradient clipping, noise calibration, patient-grouped sampling, secure aggregation, and complete privacy accounting. In addition, the current classification based on the level of frames does not explicitly model the temporal relationships between successive capsule-endoscopy images.

Future studies should be based on reproducible experimental validation as well as thorough verification of privacy assurances at patient level and validation by multiple independent clinical institutions. This also warrants further studies of the temporal modelling, robustness to distribution shifts and performance metrics across heterogeneous patient populations. These directions will shed light on the potential of FedCapsule to offer dependable, privacy-enhancing, clinically relevant assistance with the detection of gastrointestinal diseases.

In conclusion, FedCapsule provides a framework for investigating privacy-preserving collaborative learning for capsule-endoscopy image classification. Before it is fully validated for its generalisability, privacy assurances and clinical applicability, it cannot be considered suitable for real-world clinical deployment.

REFERENCE:

[1] S. A. El-Ghany, M. A. Mahmood, and A. A. Abd El-Aziz, “An Accurate Deep Learning-Based Computer-Aided Diagnosis System for Gastrointestinal Disease Detection Using Wireless Capsule Endoscopy Image Analysis,” *Appl. Sci.*, vol. 14, no. 22, p. 10243, Nov. 2024, doi: 10.3390/app142210243.

[2] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, “Communication-Efficient Learning of Deep Networks from Decentralized Data,” 2016, doi: 10.48550/ARXIV.1602.05629.

[3] M. J. Mascarenhas Saraiva *et al.*, “Deep learning and capsule endoscopy: automatic identification and differentiation of small bowel lesions with distinct haemorrhagic potential using a convolutional neural network,” *BMJ Open Gastroenterol.*, vol. 8, no. 1, p. e000753, Sep. 2021, doi: 10.1136/bmjgast-2021-000753.

[4] M. Abadi *et al.*, “Deep Learning with Differential Privacy,” in *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, Vienna Austria: ACM, Oct. 2016, pp. 308–318. doi: 10.1145/2976749.2978318.

[5] M. Mohammadi *et al.*, "Differential privacy for medical deep learning: methods, tradeoffs, and deployment implications," *Npj Digit. Med.*, vol. 9, no. 1, p. 93, Jan. 2026, doi: 10.1038/s41746-025-02280-z.

[6] M. Oquab *et al.*, "DINOv2: Learning Robust Visual Features without Supervision," 2023, *arXiv*. doi: 10.48550/ARXIV.2304.07193.

[7] T. R. Kumar, S. R. Vinta, S. D. Pande, and A. Khamparia, *AI and Machine Learning for Mechanical and Electrical Engineering*, 1st ed. Boca Raton: Auerbach Publications, 2025. doi: 10.1201/9781003479062.

[8] M. Adnan, S. Kalra, J. C. Cresswell, G. W. Taylor, and H. R. Tizhoosh, "Federated learning and differential privacy for medical image analysis," *Sci. Rep.*, vol. 12, no. 1, p. 1953, Feb. 2022, doi: 10.1038/s41598-022-05539-7.

[9] Bhau, Gaikar Vilas, et al. "IoT based solar energy monitoring system." *Materials Today: Proceedings* 80 (2023): 3697-3701.

[10] P. H. Smedsrud *et al.*, "Kvasir-Capsule, a video capsule endoscopy dataset," *Sci. Data*, vol. 8, no. 1, p. 142, May 2021, doi: 10.1038/s41597-021-00920-z.

[11] A. K. Menon, S. Jayasumana, A. S. Rawat, H. Jain, A. Veit, and S. Kumar, "Long-tail learning via logit adjustment," 2020, *arXiv*. doi: 10.48550/ARXIV.2007.07314.

[12] A. Ziller, D. Usynin, R. Braren, M. Makowski, D. Rueckert, and G. Kaissis, "Medical imaging deep learning with differential privacy," *Sci. Rep.*, vol. 11, no. 1, p. 13524, Jun. 2021, doi: 10.1038/s41598-021-93030-0.

[13] Patro, Pramoda, et al. "A hybrid approach estimates the real-time health state of a bearing by accelerated degradation tests, Machine learning." *2021 Second International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE)*. IEEE, 2021.

[14] S. Pati *et al.*, "Privacy preservation for federated learning in health care," *Patterns*, vol. 5, no. 7, p. 100974, Jul. 2024, doi: 10.1016/j.patter.2024.100974.

[15] S. P. Karimireddy, S. Kale, M. Mohri, S. J. Reddi, S. U. Stich, and A. T. Suresh, "SCAFFOLD: Stochastic Controlled Averaging for Federated Learning," 2019, *arXiv*. doi: 10.48550/ARXIV.1910.06378.

[16] S. A. Alghamdi *et al.*, "Safety Survey on Lone Working Magnetic Resonance Imaging Technologists in Saudi Arabia," *Healthcare*, vol. 11, no. 5, p. 721, Mar. 2023, doi: 10.3390/healthcare11050721.

[17] T. T. Habe, K. Haataja, and P. Toivanen, "Review of Deep Learning Performance in Wireless Capsule Endoscopy Images for GI Disease Classification," *F1000Research*, vol. 13, p. 201, Mar. 2024, doi: 10.12688/f1000research.145950.1.

[18] I. Mironov, "Rényi Differential Privacy," in *2017 IEEE 30th Computer Security Foundations Symposium (CSF)*, Santa Barbara, CA: IEEE, Aug. 2017, pp. 263–275. doi: 10.1109/CSF.2017.11.

[19] A. Sadilek *et al.*, "Privacy-first health research with federated learning," *Npj Digit. Med.*, vol. 4, no. 1, p. 132, Sep. 2021, doi: 10.1038/s41746-021-00489-2.