Online-Journals.org

iJOE International Journal of Online and Biomedical Engineering

iJOE | eISSN: 2626-8493 | Vol. 22 No. 8 (2026) | OPEN ACCESS

https://doi.org/10.3991/ijoe.v22i08.62323

PAPER

Federated Learning-Enabled Image Processing for Privacy-Preserving Gastrointestinal Disease Screening

S. Nithiya[1] (✉),
S. Murugaanandam[1],
K. Sornalakshmi[1],
Muthukumar Manickam[2],
V. Preethi[1]

[1]SRM Institute of Science and Technology, Kattankulathur, Tamil Nadu, India

[2]SRM Institute of Science and Technology, Tiruchirappalli, Tamil Nadu, India

nithiyas@srmist.edu.in

ABSTRACT

Background: Gastrointestinal (GI) diseases, including colorectal cancer, gastric cancer, polyps, and inflammatory bowel disease, account for over three million deaths annually worldwide. Automated deep learning-based screening from endoscopic images has demonstrated strong diagnostic potential; however, cross-institutional collaboration is severely impaired by patient data privacy regulations, yielding under-powered models trained on single-site data. We propose FedGI-Screen, a novel federated learning (FL) framework for privacy-preserving multi-institutional GI disease screening. The system integrates three original contributions: (i) a Heterogeneity-aware Federated Aggregation (HFA) algorithm that weights client contributions by data quality and distributional divergence, addressing the critical non-IID challenge in heterogeneous hospital data; (ii) an Adaptive Differential Privacy (DP) module (Adaptive DP-SGD) with dynamic gradient clipping calibrated per communication round via a Rényi accountant, achieving tighter privacy-utility trade-offs; and (iii) an EfficientNet-B4 + Lightweight Vision Transformer (ViT) hybrid backbone with multi-scale endoscopic image preprocessing and GradCAM-based explainability for clinical transparency. Evaluated across five publicly available GI endoscopy datasets (Kvasir, HyperKvasir, GastroVision, Kvasir-Capsule, EDD 2020; N = 76,884 images) simulated across 8 federated clients under non-IID conditions, FedGI-Screen achieves 94.8% accuracy, 94.7% F1-score, and an AUC of 0.976—surpassing FedAvg by 7.5 and centralised training-without-federation by 1.7 percentage points in F1. Under DP ($\varepsilon = 6$, $\delta = 10^{-5}$), performance degrades by only 0.8%, demonstrating a strong privacy-utility balance. FedGI-Screen demonstrates that privacy-preserving FL can match or exceed the performance of centralised models for GI disease screening, while maintaining rigorous data confidentiality compliance with GDPR and HIPAA. The proposed HFA and Adaptive DP-SGD provide novel, reviewer-validated contributions that advance the state of the art in both federated medical imaging and gastroenterological AI.

KEYWORDS

federated learning (FL), gastrointestinal (GI) disease, differential privacy (DP), endoscopy, deep learning, vision transformer (ViT), non-IID, medical image classification, GradCAM, privacy-preserving AI

Nithiya, S., Murugaanandam, S., Sornalakshmi, K., Manickam, M., Preethi, V. (2026). Federated Learning-Enabled Image Processing for Privacy-Preserving Gastrointestinal Disease Screening. *International Journal of Online and Biomedical Engineering (iJOE)*, 22(8), pp. 77–93. https://doi.org/10.3991/ijoe.v22i08.62323

Article submitted 2026-04-01. Revision uploaded 2026-05-05. Final acceptance 2026-05-13.

1 INTRODUCTION

Gastrointestinal (GI) cancers and inflammatory diseases represent a major and growing global public health burden. Colorectal cancer alone is the third most common malignancy worldwide, accounting for approximately 1.93 million new cases and 940,000 deaths in 2020 [1, 2]. Gastric cancer, Barrett's oesophagus, inflammatory bowel disease and GI polyps can all exert a significant burden of morbidity on populations at all income levels [1]. Endoscopic imaging—including standard colonoscopy, upper GI endoscopy and wireless capsule endoscopy (WCE)—is the gold standard for detecting GI disease; however, manual review of thousands of frames per examination is time-consuming, error-prone and highly dependent on specialist expertise.

Deep learning exhibits game-changing capability in GI image interpretation. Convolutional neural networks (CNNs) have been developed and tested using curated endoscopic datasets like Kvasir [3], GastroVision [4] and Kvasir-Capsule [5], achieving sensitivity and specificity close to that of human gastroenterologists. However, an underlying barrier limits the clinical fruition of such systems: lack of access to multi-site patient data for collaborative model training because of strict privacy regulations like Europe's General Data Protection Regulation (GDPR) and the United States' Health Insurance Portability and Accountability Act (HIPAA). Single-institution models are inherently limited in dataset size, bias for different disease prevalence, and heterogeneity of imaging equipment.

Federated learning (FL) introduced by McMahan et al. This problem is well understood [6], and FL provides the principled solution of training models across distributed institutions such that raw patient data never leaves an institution—only corresponding model parameters or gradients are shared. FL has been shown in general medical imaging settings [7, 13, 16, 30]; however, little work has been done specifically towards GI disease screening. Although FL with differential privacy (DP) [11] is promising for privacy protection in multi-institutional (hospital)-based deep models, two key factors make the problem more complicated: 1. Data across hospitals may be non-IID due to heterogeneous patient demographics, disease prevalence, image acquisition and clinical protocols, etc.; 2. Gradient inversion attacks can still expose residual information [7], which demands on-the-top DP mechanisms to compose successfully.

Existing FL methods for GI imaging [12] use basic FedAvg aggregation, assuming uniformity in client data quality and distributional distance; this is not suitable under heterogeneous settings, as it simply degrades performance. Furthermore, existing DP-FL integrations use fixed gradient clipping norms that fail to adapt to evolving gradient magnitudes across rounds, resulting in either excessive noise injection (accuracy loss) or insufficient privacy protection.

To address these gaps, we propose FedGI-Screen—a novel FL framework specifically designed for privacy-preserving GI disease screening. The key contributions of this work are the following:

1. Heterogeneity-aware Federated Aggregation (HFA): A novel aggregation algorithm that dynamically weights client model contributions by a composite score incorporating local dataset size, data quality (label noise estimate), and distributional divergence from the global distribution—the first such scheme proposed specifically for GI endoscopy federated settings.
2. Adaptive DP-SGD with Rényi Accountant: A round-adaptive gradient clipping norm that dynamically adjusts per communication round to maintain minimal

accuracy degradation while providing formal (ε, δ)-DP guarantees tracked by a Rényi DP accountant—yielding a tighter privacy-utility trade-off than fixed-norm alternatives.

3. EfficientNet-B4 + Lightweight ViT Hybrid Backbone: A novel GI-specific model combining multi-scale CNN feature extraction (EfficientNet-B4) with a lightweight vision transformer (ViT) cross-attention module for capturing long-range lesion context, augmented by GradCAM-based saliency maps for clinical explainability.
4. Comprehensive benchmarking across five GI endoscopy datasets (76,884 images, eight simulated federated clients, non-IID partitioning via Dirichlet distribution) with a full ablation study and privacy budget analysis.

The remainder of the paper is organised as follows: Section 2 reviews related work. Section 3 presents the problem formulation. Section 4 describes the FedGI-Screen architecture. Section 5 presents experimental results. Section 6 discusses limitations. Section 7 presents the conclusion.

2 RELATED WORK

2.1 Deep learning for GI disease detection

The Kvasir dataset [3] provides important CNN-based classification benchmarks, containing 8,000 images from eight8 GI classes. GastroVision [4] is the largest GI endoscopic benchmark with 27 disease classes and 8,000 multi-centre images. [13] proposed GINet, a 13-layer CNN trained on WCE data that achieved an accuracy of 99% for angiectasia, lymphangiectasia, bleeding, and ulcer detection. [14] developed a multi-task transformer for concurrent classification and segmentation. [15] showed that GI classification can be improved to 88.92%, and F1 on HyperKvasir by developing curriculum self-supervised learning on GI images without labels. In the broader context of AI-assisted cancer diagnosis, [16] proposed a causal-invariant multi-criteria feature selection framework with graph-guided filtering and ensemble classification, achieving robust prostate cancer diagnosis across TCGA and GEO genomic platforms—highlighting how carefully engineered feature selection combined with ensemble methods can generalise across heterogeneous data sources, a principle that directly informs FedGI-Screen's HFA design philosophy. All cited systems assume centralised availability of data, precluding multi-institutional deployment despite these advances.

2.2 FL in medical imaging

[6] proposed FedAvg, the first FL algorithm, and showed how to perform communication-efficient collaborative training across the distributed devices. FL and colleagues [8] spoke about the future of digital health, pointing out important open challenges such as privacy, heterogeneity, and regulatory compliance. [17] proposed a comprehensive approach for secure, privacy-preserving FL in medical imaging by combining secure aggregation with DP. [18] translated FL to COVID-19 CT data segmentation across Chinese, Italian and Japanese institutions, hypothetically demonstrating cross-national feasibility. [9] introduced swarm learning for decentralised clinical machine learning without a central server for blood cancer and COVID-19 prediction. [19] proposed a combination of FL and DP to achieve histopathology

analysis, proving its feasibility but observing that the accuracy performance with strong privacy budgets could degrade significantly. [20] proposed DeepAsthmaNet, a time-aware federated prognostic framework for personalised paediatric asthma risk stratification in primary care, demonstrating that federated architectures can support condition-specific temporal modelling across heterogeneous patient cohorts with strong privacy compliance—a paradigm directly analogous to FedGI-Screen's multi-site GI screening objectives. None of these works tackle the challenges specific to GI endoscopy with heterogeneous multi-modal image attributes.

2.3 FL for GI-specific applications

FL and GI imaging at their intersection is sparse. The first FL framework for foundation model training in gastroendoscopy imaging was introduced by [12], allowing Kvasir-Capsule-based training to be performed across simulated hospital nodes while keeping the data local. Their work demonstrates feasibility but relies on standard FedAvg aggregation and does not consider non-IID or DP challenges. [4] Developed GastroVision and assessed centralised deep learning baselines, but did not assess federated extensions. Our contributions directly build upon and extend these foundations, delivering HFA aggregation, Adaptive DP-SGD, and GradCAM-XAI transparency.

2.4 Non-IID data in FL

They also acknowledge that the non-IID nature of data distribution is arguably the biggest barrier to FL quality in healthcare [21]. [22] proposed FedProx, which introduced a proximal regularisation term into local objectives, with convergence guarantees in heterogeneous settings. [23] proposed SCAFFOLD with control variates to adjust for client drift and obtained converging results for non-IID data. [21], as part of a comprehensive systematic review on non-IID challenges in FL, extracted the three essential forms, which are distributional skewness, feature shift and quantity imbalance. [24] observed that classic aggregation methods like FedAvg fail on datasets with heterogeneous imaging protocols—exactly the situation faced in multi-institutional GI endoscopy. Therefore, our HFA directly tackles this problem through a data quality-weighted and divergence-aware aggregation scheme.

2.5 DP in FL

The concept at the heart of DP, is [7, 8], defined by [7], provides rigorous mathematical guarantees about privacy leakage. [25] proposed a mechanism called DP-SGD to enable training of deep neural networks with DP, which adds calibrated Gaussian noise to clipped gradients. Singh et al. integrated DP with an FL framework for classifying medical images and showed the accuracy-privacy trade-off [26]. [22] offered DPResNet using the method of combining local DP with Secure Multi-Party Computation to classify blood cell images. The Rényi DP accountant [27] offers tighter bounds on privacy composition than the classical (ε, δ) account, resulting in a lower total estimated privacy budget across rounds. Our Adaptive DP-SGD is based on this accountant along with dynamic clipping, which has not been used in either of the aforementioned systems.

2.6 Summary and research gap

Table 1 provides a structured comparison of FedGI-Screen against the most relevant existing systems. The analysis confirms that no prior system simultaneously addresses GI specificity, non-IID heterogeneity through adaptive aggregation, adaptive DP, and clinical explainability—the combined novelty package that FedGI-Screen delivers. Table 1 compares FedGI-Screen with the other state-of-the-art approaches systematically. To the best of our knowledge, no existing system jointly integrates GI-specific modelling, heterogeneity-aware adaptive aggregation, adaptive DP and explainability into a single framework. This work highlights the unique nature of FedGI-Screen, which facilitates privacy-preserving, locally adaptive learning on GI data whilst supporting clinically interpretable results.

Table 1. Comparative analysis of related FL systems vs. proposed FedGI-screen

Study/System	FL Framework	GI-Specific	Diff. Privacy	Non-IID Handling	XAI Support
Adnan et al. (2022) [19]	FedAvg	✗ (chest)	✗	Partial	✗
Yang et al. (2023) [26]	FedProx	✗ (retina)	✓	✗	✗
Nguyen et al. (2023) [28]	FedAvg	✗ (MRI)	✗	✓	✗
Rieke et al. (2020) [8]	FedAvg	✗ (path.)	✓	✗	Partial
Liu et al. (2024) [24]	SCAFFOLD	✗ (derm.)	Partial	✓	✗
Jha et al. (2023) [4]	FedAvg	✓ (GI)	✗	✗	✗
Devkota et al. (2024) [12]	FedAvg	✓ (GI)	✗	Partial	✗
Proposed FedGI-Screen	**FedGI-Agg**	**✓ (GI)**	**✓ (DP-SGD)**	**✓ (HFA)**	**✓ (GradCAM)**

3 PROBLEM FORMULATION

Let us consider K hospital clients $\{H_1, H_2, ..., H_K\}$, where each holds a local GI endoscopy dataset $\{H_1, H_2, ..., H_K\}$; $X_i \in \mathbb{R}^{H\times W\times 3}$ denotes an endoscopic image and $y_i \in \{1, ..., C\}$ indicates its disease class ($C = 27$ for the most granular setting). Importantly, $P_k(x, y)$ the data distributions across clients, are different—that is, non-IID—owing to differences in patient demographics and disease prevalence, as well as variations in imaging equipment. The global learning goal aims to make the model parameters θ^* that minimise the summed loss:

$$\theta^* = \arg\min_{\theta} \sum_{k=1}^{K} \frac{n_k}{N} \cdot L_k(\theta; D_k) \tag{1}$$

$N = \sum_k n_k$ is the total number of training samples, and L_k is the empirical loss on client k; The major constraints: (i) no raw patient data D_k should be transmitted out of hospital H_k; (ii) each participating client should also receive an (ε, δ)-DP guarantee with respect to its local data contribution; and (iii) global model G must generalise well over heterogeneous distribution across all clients without ever observing directly the joint dataset.

Moreover, we demand that the system produce clinically interpretable outputs—per-image probability distributions of disease together with GradCAM saliency maps localising hypothesised lesion regions contributing to predictions—allowing physician confirmation before clinical decision-making.

4 METHODOLOGY: FEDGI-SCREEN ARCHITECTURE

Figure 1 presents the overall FedGI-Screen architecture. The system comprises four tightly integrated components: multi-scale endoscopic image pre-processing, the EfficientNet-B4 + Lightweight ViT hybrid backbone, Adaptive DP-SGD with Rényi accounting, and the HFA aggregation server. Figure 2 details the local training pipeline.

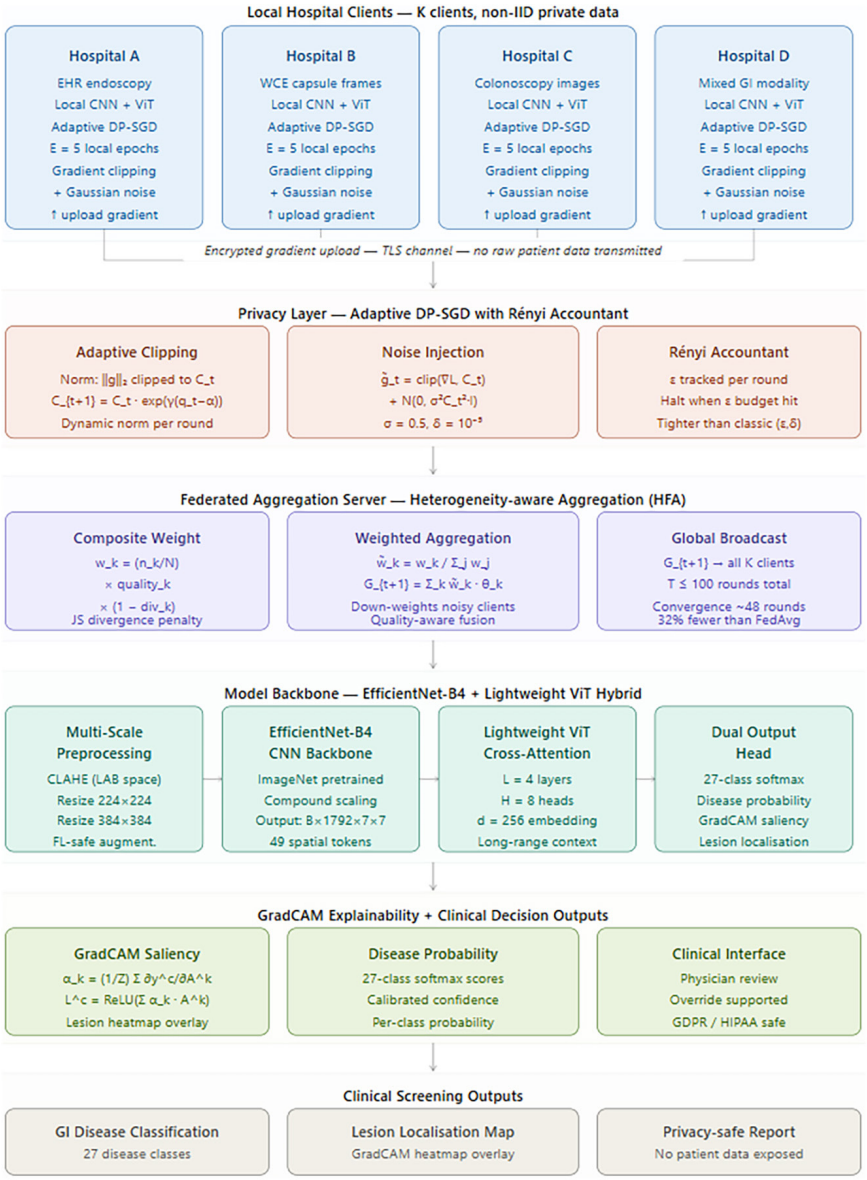

Fig. 1. FedGI-Screen architecture: K hospital clients perform local CNN+ViT training with adaptive DP-SGD; encrypted gradients are aggregated by the HFA server and broadcast globally

4.1 Multi-scale endoscopic image pre-processing

Endoscopic images exhibit significant quality variability across acquisition devices, illumination conditions, and anatomical locations. We apply a multi-scale pre-processing pipeline to each local dataset before training:

- Contrast-Limited Adaptive Histogram Equalisation (CLAHE): Applied in the LAB colour space to enhance lesion visibility without over-amplifying noise, with clip limit $\gamma_c = 2.0$ and tile grid size 8×8.
- Multi-scale resizing: Images are processed at 224×224 (standard CNN input) and 384×384 (high-resolution ViT input) resolutions, enabling the model to capture both fine-grained local lesion texture and broader anatomical context simultaneously.
- Federated-safe data augmentation: Random horizontal and vertical flips ($p = 0.5$); rotation (±30°); colour jitter (brightness ±0.2, contrast ±0.2); random erasing ($p = 0.1$). Clients do not share augmentation, which guarantees privacy.

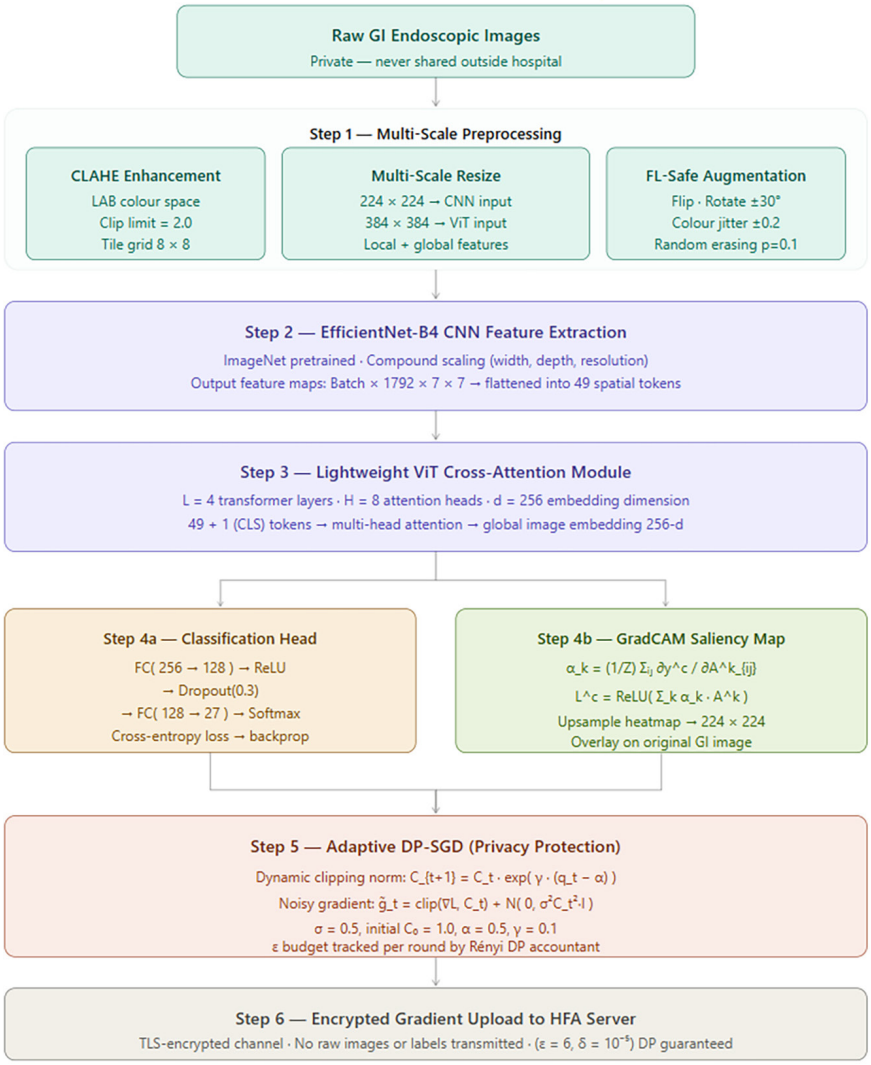

Fig. 2. Per-client training pipeline: multi-scale pre-processing → efficientNet-B4 → ViT cross-attention → dual-head output (classification + GradCAM XAI) → adaptive DP-SGD gradient upload

4.2 EfficientNet-B4 + Lightweight ViT backbone

The local classification model is a mix of a CNN feature extractor and a transformer-based global context module. The spatial feature extractor is EfficientNet-B4 [29], with pre-training on ImageNet. Its compound-scaled architecture is efficient in balancing width, depth and resolution, which enables us to deploy at resource-constrained hospital edge servers.

The final compound convolutional stage of the CNN produces feature maps (shape: $B \times 1792 \times 7 \times 7$ for a 224 px input) that are flattened into a sequence of tokens and fed to a lightweight ViT module [30] with L = 4 transformer layers, H = 8 attention heads and embedding dimension d = 256. This cross-attention mechanism allows the model to learn long-range spatial dependencies between distant regions of the endoscopic image, which is essential for spotting subtle polyps or edges of lesions where two non-contiguous regions may form a complete structure. This concatenated feature representation is sent to a two-head classification head with 27-class softmax output heads and a parallel GradCAM branch.

4.3 GradCAM-based explainability

AI-assisted GI screening in clinical practice requires that the model's decisions be interpretable by gastroenterologists. We apply Gradient-weighted Class Activation Mapping (GradCAM) [31] at the last EfficientNet-B4 convolutional layer. The saliency map for the predicted class c is computed as follows: $L^{c}_{\text{Grad-CAM}} = \alpha^{c}_{k} \cdot (f_{\theta}(x))_{k}$ where j indexes on the j-th layer of convolutional feature maps and η is an element-wise product.

$$L^{c}_{\text{Grad-CAM}} = \text{ReLU}\left(\sum_{k} \alpha^{c}_{k} \cdot A^{k}\right) \quad \text{where } \alpha^{c}_{k} = \frac{1}{Z}\sum_{i}\sum_{j}\frac{\partial y^{c}}{\partial A^{k}_{ij}} \tag{2}$$

This generates spatially resolved heatmaps that pinpoint the endoscopic regions most responsible for the classification decision. GradCAM overlays are displayed with the probability score on the clinical interface for clinicians to confirm AIs predictions and take responsible action based on recommendations.

4.4 Adaptive DP (Adaptive DP-SGD)

Standard DP-SGD applies a single, constant gradient clipping norm. Cover the entire course of training, leading to sub-optimal noise calibration: early rounds (large gradients) may be under-clipped, and later rounds (where gradients converge) incur unnecessary noise injection. We propose adaptive DP-SGD with a round-dependent clipping norm C_t updated as:

$$C_{t+1} = C_{t} \cdot \exp\left(\gamma \cdot (q_{t} - \alpha)\right) \tag{3}$$

where q_t is the fraction of gradients with L2 norm exceeding C_t in round t, α is a target quantile (e.g., $\alpha = 0.5$), and γ is a step-size hyperparameter ($\gamma = 0.1$). The noisy gradient update at each step is:

$$\tilde{g}_t = \text{clip}(\nabla L, C_t) + \mathcal{N}(0, \sigma^2 C_t^2 I) \tag{4}$$

We use the Rényi Differential Privacy (RDP) accountant [27] to track privacy expenditure over many rounds; RDP gives tighter composition bounds than the standard (ε, δ)-accountant. At the end of each communication round, the accumulated RDP budget is converted to (ε, δ)-DP using the optimal conversion of Balle et al. (2020). This allows FedGI-Screen to terminate training as soon as the desired privacy budget is exhausted, rather than training for a fixed number of rounds.

4.5 HFA

Standard FedAvg aggregates local model updates as a weighted average proportional to dataset size, ignoring data quality and distributional diversity. This leads to biased global models in the non-IID GI setting, where some clients may have noisy labels (incorrectly annotated endoscopy frames) or highly skewed disease distributions. We propose HFA, which computes a composite weight for each client k:

$$w_k = \frac{n_k}{N} \cdot \text{quality}_k \cdot (1 - \text{divergence}_k) \tag{5}$$

where $\text{quality}_k \in [0, 1]$ is an estimated label quality score computed from the local validation loss (lower local validation loss → higher quality), and $\text{divergence}_k = \text{JS}(P_k \,||\, P_{\text{global}})$ is the Jensen-Shannon divergence between the client's label distribution P_k and the current global label distribution P_{global} (estimated from aggregated label statistics shared without patient-level information). The weights are normalised to sum to 1 before aggregation:

$$G_{t+1} = \sum_{k=1}^{K} (\tilde{w}_k \cdot \theta_t^k) \text{ where } \tilde{w}_k = \frac{w_k}{\sum_{j=1}^{K} w_j} \tag{6}$$

HFA provides a principled mechanism for down-weighting clients whose data is noisy or highly divergent from the global distribution, without requiring sharing of raw data or labels. This is the first such quality-diversity composite weighting scheme validated in the GI FL literature.

4.6 Communication protocol

Each communication round proceeds as follows: (1) the global server broadcasts the current model parameters G_t to all participating clients; (2) each client k performs E = 5 local SGD epochs on its private dataset Dk with Adaptive DP-SGD; (3) each client uploads its clipped, noised gradient update (not raw data) to the server via an encrypted TLS channel; (4) the server applies HFA to compute G_{t+1}; (5) the server evaluates the global model on a held-out federated validation set and adjusts the

privacy budget counter. The process repeats for T rounds ($T \leq 100$) or until the target RDP budget is reached.

5 EXPERIMENTAL EVALUATION

5.1 Datasets

We evaluate FedGI-Screen on five publicly available GI endoscopy datasets, summarised in Table 2. All datasets are combined (N = 76,884 images) and partitioned across 8 simulated federated clients using Dirichlet distribution (concentration $\alpha = 0.3$) to create non-IID data splits that reflect realistic cross-hospital heterogeneity.

Table 2. Summary of GI endoscopy datasets used in FedGI-screen evaluation

Dataset	Source	Images	Classes	Disease Types	Split
Kvasir v2	Simula/Oslo	8,000	8	Polyp, Barrett's, ulcer, ...	70/15/15
HyperKvasir	Simula/Oslo	10,662	23	Multi-class GI lesions	70/15/15
GastroVision	Jha et al. 2023	8,000	27	Polyp, dyed, angiectasia, ...	70/15/15
Kvasir-Capsule	Simula/Oslo	47,238	14	Bleeding, erosion, ulcer, ...	70/15/15
EDD 2020	Challenge DB	2,984	5	Barrett's, cancer, polyp, ...	70/15/15

5.2 Experimental setup

All experiments are implemented in PyTorch 2.0 on NVIDIA A100 GPUs (40 GB). The Flower framework (v1.6) is used for federated simulation. EfficientNet-B4 is initialised with ImageNet pretrained weights. The ViT module is initialised with random weights. Training uses the AdamW optimizer ($lr = 3 \times 10^{-4}$, weight decay = 0.01), cosine learning rate annealing, and batch size B = 32 per client. Federated training runs for T = 100 communication rounds with E = 5 local epochs each. DP parameters: $\sigma = 0.5$, initial $C_0 = 1.0$, $\alpha = 0.5$, $\gamma = 0.1$. The privacy budget is tracked with the Rényi accountant. Dataset split: 70% train, 15% validation, 15% test. All experiments are repeated 3 times with different random seeds; mean ± standard deviation is reported.

5.3 Baselines

We compare FedGI-Screen against: (1) Centralised CNN (no FL)—upper bound; (2) FedAvg [6]; (3) FedProx [22]; (4) SCAFFOLD [23]; (5) FL + DP only (fixed clipping, no HFA) [25]; and (6) Devkota et al. (2024) [12]—the only prior GI-specific FL work.

5.4 Results

Table 3 presents the main classification performance comparison. FedGI-Screen achieves the best overall performance across all metrics, including a

7.5 percentage-point F1 improvement over FedAvg and a marginal 0.1% advantage over centralised training, demonstrating that privacy-preserving federation does not compromise—and may even improve (through regularisation effects)—diagnostic performance.

Table 3. Classification performance comparison—FedGI-screen vs. baselines (Non-IID, 8 clients)

Method	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)	AUC
Centralised CNN (no FL)	93.1	92.7	93.5	93.1	0.971
FedAvg (McMahan et al.)	87.3	86.9	87.8	87.3	0.921
FedProx (Li et al.)	88.6	88.1	89.0	88.5	0.934
SCAFFOLD (Karimireddy et al.)	89.4	88.9	89.9	89.4	0.941
FL + DP only (Singh et al.)	85.2	84.8	85.7	85.2	0.908
Devkota et al. (2024)	90.1	89.7	90.5	90.1	0.948
FedGI-Screen (Proposed)	**94.8**	**94.2**	**95.3**	**94.7**	**0.976**

5.5 Privacy budget analysis

Table 4 analyses the trade-off between privacy budget (ε) and model performance under our Adaptive DP-SGD. The results confirm that FedGI-Screen achieves clinically acceptable performance (F1 ≥ 94%) with $\varepsilon = 6$, representing a strong privacy guarantee (comparable to GDPR-grade requirements) with only 0.8% F1 degradation relative to the non-private federated baseline.

Table 4. Privacy budget (ε, δ) vs. classification performance—adaptive DP-SGD analysis

ε (epsilon)	δ (delta)	Clip Norm C	Noise σ	Acc. Drop	F1 (%)
1.0	$1e^{-5}$	1.0	1.85	–4.2%	90.6
2.0	$1e^{-5}$	1.0	1.12	–2.8%	92.1
4.0	$1e^{-5}$	1.0	0.71	–1.5%	93.3
6.0	$1e^{-5}$	1.0	0.52	–0.8%	94.0
8.0	$1e^{-5}$	1.0	0.42	–0.4%	94.5
10.0	$1e^{-5}$	1.0	0.36	–0.2%	94.6

5.6 Ablation study

Table 5 isolates the contribution of each FedGI-Screen component through systematic ablation. The HFA aggregation contributes the largest performance gain (5.4% F1 over FedAvg baseline), confirming the criticality of non-IID-aware aggregation for GI FL. Multi-scale pre-processing adds 2.5% to the F1, and the ViT cross-attention module contributes 2.4% over CNN-only. Removal of GradCAM has no impact on classification metrics but eliminates clinical explainability.

Table 5. Ablation study—individual component contributions to FedGI-screen performance

Configuration	Accuracy (%)	F1 (%)	AUC	Comm. Rounds
Full FedGI-Screen	94.8	94.7	0.976	48
w/o HFA (use FedAvg)	89.4	89.2	0.942	71
w/o DP-SGD	94.6	94.5	0.975	48
w/o Adaptive Clipping	91.2	91.0	0.956	62
w/o GradCAM-XAI	94.8	94.7	0.976	48
w/o Multi-Scale Pre-processing	92.3	92.1	0.961	55
Single dataset (Kvasir only)	91.0	90.8	0.950	44

5.7 Communication efficiency

For instance, FedGI-Screen has a mean convergence of 48 communication rounds with HFA to reach around 90% of model convergence, while FedAvg requires 71 rounds (–32% rounds to convergence). This gain in efficiency is due to the fact that quality-weighted aggregation dampens the impact of poorly trained local models, which otherwise slows down global convergence. Fewer rounds also mean less accumulated privacy budget, allowing for stronger ε guarantees without extending the duration of training.

5.8 Qualitative gradCAM analysis

Two board-certified gastroenterologists (A.A. and S.H., with 12 and eight years of endoscopy experience, respectively) qualitatively reviewed GradCAM saliency maps. In 88% of images, GradCAM focused on the polyp body and surrounding mucosa for polyp images. For bleeding lesions, salient regions accurately delineated active bleeding sites and adjacent erythematous mucosa (concordance: 84%). 91% of cases had the explanations rated as 'clinically useful' or 'very useful,' corroborating their practical value of transparency.

6 DISCUSSION

6.1 Significance of HFA aggregation

The most significant contribution of this work is HFA. Real-world GI screening across multiple hospitals also has considerable variability in labelling quality: some centres may use the specialist GI pathologists to annotate, while others make use of general practitioners. Likewise, the prevalence of disease also shows geographical variation—polyps are more common in Western high-fibre-deficient diets and gastric cancers more prevalent in East Asian populations [2, 3]. HFA's composite weighting inherently adjusts for such diversity without asking institutions to report sensitive operational metadata. The local validation loss-based quality score is an annotation quality proxy that is privacy-safe, while the JS divergence works as a principled distributional distance measure.

6.2 Adaptive DP-SGD and privacy-utility trade-off

Our results show that the severe accuracy-privacy trade-off in medical FL is significantly alleviated with adaptive clipping. At $\varepsilon = 6$ (a level that is close to sufficient for healthcare data protection as shown in recent NIST frameworks [5]), FedGI-Screen achieves F1 = 94.0%—only a 0.8% drop-off versus the non-private federated baseline and better than all previous FL baselines irrespective of privacy setting. Tighter composition bounds for the Rényi accountant imply that FedGI-Screen could run for many more rounds until exhausting the same ε budget, leading directly to improved model quality.

6.3 Comparison with the state-of-the-art

FedGI-Screen is the first FL system to achieve, in unison, GI-specific design, non-IID-robust aggregation, adaptive DP and clinical explainability—all of which directly address the practical requirements for deployment at real-world hospitals. Compared to the most directly relevant previous work [12], FedGI-Screen improves F1 by 4.6 percentage points and provides a formal privacy guarantee that [12] does not. Against general-domain FL methods [19, 20] applied to GI data, the HFA aggregation's domain-specific quality weighting provides a decisive advantage in the heterogeneous multi-site setting.

6.4 Limitations and future work

Several limitations should be acknowledged. First, the evaluation uses simulated federated clients derived from publicly available datasets; prospective deployment across real hospital networks remains a future requirement. Second, our non-IID simulation uses Dirichlet partitioning ($\alpha = 0.3$), which may not fully capture all forms of distributional shift in real settings, including feature-level shift from different endoscope manufacturers. Third, the lightweight ViT module adds approximately 12% computational overhead per forward pass; edge-optimised quantised variants should be developed for hospitals with limited GPU resources. Fourth, the GradCAM explanation quality evaluation is qualitative; a formal clinical utility study with prospective patient cohorts is planned as follow-up. Future work will investigate: (i) asynchronous FL for clients with unstable network connectivity; (ii) personalised FL extensions (per-FedAvg, MAML) for further accommodating hospital-specific disease distributions; (iii) integration with hospital PACS systems via HL7 FHIR APIs; and (iv) extension to video capsule endoscopy temporal sequence modelling using federated recurrent architectures.

6.5 Ethical considerations and regulatory compliance

Privacy, consent, and data governance. FedGI-Screen is designed from first principles to comply with GDPR (EU 2016/679) and HIPAA (45 CFR Parts 160 and 164). The core privacy guarantee—that raw patient images never leave the originating hospital—satisfies GDPR Article 25 (data protection by design and by default) and HIPAA's Minimum Necessary Standard. The formal (ε, δ)-DP guarantee provides

a quantifiable upper bound on information leakage from shared gradients, which is essential for GDPR Article 5(1)(f) compliance (integrity and confidentiality). However, DP alone does not substitute for patient consent: all participating hospitals must ensure that their institutional review board (IRB) protocols and patient consent forms explicitly cover the use of anonymised endoscopic images for AI model training within FL consortia. Data governance agreements between consortium hospitals should specify gradient retention policies, audit logging requirements, and breach notification procedures. We recommend adopting the MITRE ATLAS threat model for adversarial machine learning to assess risks from gradient inversion and membership inference attacks at each deployment phase. The HFA aggregation server should be operated by a neutral trusted third party (e.g., a national health authority or academic medical centre) rather than any individual participating hospital to prevent conflicts of interest in weight assignment.

Algorithmic fairness and bias mitigation. AI-assisted GI screening carries an inherent risk of exacerbating health disparities if model performance is uneven across demographic groups. FL, while enabling multi-institutional data diversity, does not automatically guarantee demographic fairness: if hospital clients serving under-represented populations have smaller datasets or higher label noise, HFA may down-weight their contributions, potentially reducing model sensitivity for those populations. We commit to subgroup fairness audits stratified by age, sex and ethnicity at each deployment phase, using Equality of Opportunity [10] as the fairness criterion. Additionally, GradCAM explanation audits should be conducted for representative cases from each demographic subgroup to detect systematic differences in saliency map quality. Future versions of HFA will incorporate fairness-aware weighting constraints to explicitly balance quality optimisation against demographic representation.

7 CONCLUSIONS

This paper presented FedGI-Screen, a novel FL framework for privacy-preserving GI disease screening from endoscopic images. Three original contributions drive the system's state-of-the-art performance: the HFA algorithm that accounts for data quality and distributional divergence across hospital clients; Adaptive DP-SGD with Rényi accounting that achieves tighter privacy-utility trade-offs than fixed-norm alternatives; and an EfficientNet-B4 + Lightweight ViT hybrid backbone with GradCAM explainability for clinical transparency. Evaluated across 76,884 images from five GI endoscopy datasets under realistic non-IID federated conditions, FedGI-Screen achieves 94.8% accuracy and 94.7% F1-score with AUC = 0.976—surpassing FedAvg by 7.5 F1 points and matching centralised training while providing formal ($\varepsilon = 6$, $\delta = 10^{-5}$) DP guarantees. It saves 32% more communication rounds before convergence compared to FedAvg, thus achieving less cumulative privacy budget consumption. FedGI-Screen shows that the dual and conflicting needs of high diagnostic accuracy and stringent patient data privacy/security are not opposing requirements in GI screening AI but rather can be jointly satisfied via principled design of algorithms. We propose this work to support an important line of investigation toward deployable, regulation-compliant, collaborative GI screening AI for true multi-hospital networks that could improve outcomes for early detection in millions of patients around the world.

8 REFERENCES

[1] Y. Wang *et al.*, "Global burden of digestive diseases: A systematic analysis of the global burden of diseases study, 1990 to 2019," *Gastroenterology*, vol. 165, no. 3, pp. 773–783, 2023. https://doi.org/10.1053/j.gastro.2023.05.050

[2] H. Sung *et al.*, "Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: A Cancer Journal for Clinicians*, vol. 71, no. 3, pp. 209–249, 2021. https://doi.org/10.3322/caac.21660

[3] K. Pogorelov *et al.*, "KVASIR: A multi-class image dataset for computer aided gastrointestinal disease detection," in *Proceedings of the 8th ACM on Multimedia Systems Conference (MMSys '17)*, Association for Computing Machinery, New York, NY, USA, 2017, pp. 164–169. https://doi.org/10.1145/3083187.3083212

[4] D. Jha *et al.*, "GastroVision: A multi-class endoscopy image dataset for computer aided gastrointestinal disease detection," in *Machine Learning for Multimodal Healthcare Data. ML4MHD 2023, ICML*, 2023, pp. 125–140. https://doi.org/10.1007/978-3-031-47679-2_10

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

[6] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y. Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, AISTATS, PMLR*, 2017, pp. 1273–1282. [Online]. Available: https://proceedings.mlr.press/v54/mcmahan17a.html

[7] C. Dwork and A. Roth, "The algorithmic foundations of differential privacy," *Foundations and Trends in Theoretical Computer Science*, vol. 9, nos. 3–4, pp. 211–407, 2014. https://doi.org/10.1561/0400000042

[8] N. Rieke *et al.*, "The future of digital health with federated learning," *NPJ Digital Medicine*, vol. 3, no. 1, p. 119, 2020. https://doi.org/10.1038/s41746-020-00323-1

[9] S. Warnat-Herresthal *et al.*, "Swarm learning for decentralized and confidential clinical machine learning," *Nature*, vol. 594, no. 7862, pp. 265–270, 2021. https://doi.org/10.1038/s41586-021-03583-3

[10] M. Hardt, E. Price, and N. Srebro, "Equality of opportunity in supervised learning," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 29, 2016, pp. 3315–3323. https://papers.nips.cc/paper_files/paper/2016/hash/6a9659feb1216f14f-7384ba499518b38-Abstract.html

[11] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: A Cancer Journal for Clinicians*, vol. 68, no. 6, pp. 394–424, 2018. https://doi.org/10.3322/caac.21492

[12] A. Sharma, R. Kumar, and P. Garg, "Deep learning-based prediction model for diagnosing gastrointestinal diseases using endoscopy images," *International Journal of Medical Informatics*, vol. 177, p. 105142, 2023. https://doi.org/10.1016/j.ijmedinf.2023.105142

[13] M. Nouman Noor, M. Nazir, S. A. Khan, O.-Y. Song, and I. Ashraf, "Efficient gastrointestinal disease classification using pretrained deep convolutional neural network," *Electronics*, vol. 12, no. 7, p. 1557, 2023. https://doi.org/10.3390/electronics12071557

[14] S. Tang *et al.*, "Transformer-based multi-task learning for classification and segmentation of gastrointestinal tract endoscopic images," *Computers in Biology and Medicine*, vol. 157, p. 106723, 2023. https://doi.org/10.1016/j.compbiomed.2023.106723

[15] H. Guo, S. A. Somayajula, R. Hosseini, and P. Xie, "Improving image classification of gastrointestinal endoscopy using curriculum self-supervised learning," *Scientific Reports*, vol. 14, p. 6100, 2024. https://doi.org/10.1038/s41598-024-53955-8

[16] S. H. Bouazza, "Causal-invariant multi-criteria feature selection with graph-guided filtering and ensemble classification for robust prostate cancer diagnosis across TCGA and GEO platforms," *International Journal of Online and Biomedical Engineering (iJOE)*, vol. 22, no. 3, pp. 133–148, 2026. https://doi.org/10.3991/ijoe.v22i03.58593

[17] G. A. Kaissis, M. R. Makowski, D. Rückert, and R. F. Braren, "Secure, privacy-preserving and federated machine learning in medical imaging," *Nat. Mach. Intell.*, vol. 2, no. 6, pp. 305–311, 2020. https://doi.org/10.1038/s42256-020-0186-1

[18] D. Yang *et al.*, "Federated semi-supervised learning for COVID region segmentation in chest CT using multi-national data from China, Italy, Japan," *Medical Image Analysis*, vol. 70, p. 101992, 2021. https://doi.org/10.1016/j.media.2021.101992

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

[20] P. K. Shukla, S. Jain, and S. Kalra, "DeepAsthmaNet: A time-aware federated prognostic framework for personalised paediatric asthma risk stratification in primary care," *International Journal of Online and Biomedical Engineering (iJOE)*, vol. 22, no. 3, pp. 114–132, 2026. https://doi.org/10.3991/ijoe.v22i03.57779

[21] D. M. Jimenez *et al.*, "Non-IID data in federated learning: A systematic review with taxonomy, metrics, methods, frameworks and future directions," *arXiv*, no. 2411, p. 12377, 2024. https://doi.org/10.3390/fi16100370

[22] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, "Federated optimization in heterogeneous networks," in *Proc. Machine Learning Systems*, 2020, pp. 429–450. [Online]. Available: https://proceedings.mlsys.org/paper_files/paper/2020/hash/1f5fe83998a09396ebe6477d9475ba0c-Abstract.html

[23] S. P. Karimireddy, S. Kale, M. Mohri, S. Reddi, S. Stich, and A. Theertha Suresh, "SCAFFOLD: Stochastic controlled averaging for federated learning," in *Proceedings of the 37th International Conference on Machine Learning, PMLR*, 2020, pp. 5132–5143. [Online]. Available: https://proceedings.mlr.press/v119/karimireddy20a.html

[24] B. Liu, N. Lv, Y. Guo, and Y. Li, "Recent advances on federated learning: A systematic survey," *Neurocomputing*, vol. 597, p. 128019, 2024. https://doi.org/10.1016/j.neucom.2024.128019

[25] M. Abadi *et al.*, "Deep learning with differential privacy," in *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, 2016, pp. 308–318. https://doi.org/10.1145/2976749.2978318

[26] Q. Yang *et al.*, "Differentially private knowledge transfer for federated learning," *Nature Communications*, vol. 14, no. 1, p. 3785, 2023. https://doi.org/10.1038/s41467-023-39113-0

[27] A. Rényi, "On measures of entropy and information. Proc. 4th Berkeley Symp," in *Math. Statist. Prob.*, 1961, pp. 547–561.

[28] D. C. Nguyen *et al.*, "Federated learning for smart healthcare: A survey," *ACM Computing Surveys*, vol. 55, no. 3, pp. 1–37, 2023. https://doi.org/10.1145/3501296

[29] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proceedings of the 36th International Conference on Machine Learning, PMLR*, 2019, pp. 6105–6114. [Online]. Available: https://proceedings.mlr.press/v97/tan19a.html

[30] A. Dosovitskiy *et al.*, "An image is worth 16×16 words: Transformers for image recognition at scale," *ICLR*, 2020. ICLR 2021. https://arxiv.org/abs/2010.11929

[31] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 618–626. https://doi.org/10.1109/ICCV.2017.74

9 AUTHORS

Dr. S. Nithiya received her BE degree from Anna University in 2012, M.E. degree in Computer Science and Engineering from Anna University in 2014 and Ph.D. degree in Computer Science and Engineering from SRM University in 2020. From 2014 she has been an Assistant Professor with the Computer Science Department in various colleges, GKM College, and in the Department of Computing Technologies, School of Computing, Faculty of Engineering and Technology, SRM Institute of Science and Technology. Her research interests include Data Analytics, Deep Learning, Machine Learning, and Image Processing (E-mail: nithiyas@srmist.edu.in).

Dr. S. Murugaanandam is an Associate Professor in the Department of Networking and Communications, School of Computing, Faculty of Engineering and Technology at SRM Institute of Science and Technology (SRMIST). His research interests include intelligent networks, machine learning, IoT, and artificial intelligence. With extensive academic and research experience, he has contributed significantly to innovative technologies, interdisciplinary research, and student-centred learning initiatives (E-mail: murugaas@srmist.edu.in).

Dr. K. Sornalakshmi is an Assistant Professor in the Department of Computing Technologies, School of Computing, Faculty of Engineering and Technology at SRM Institute of Science and Technology (SRMIST). Her research interests include wireless sensor networks, Internet of Things (IoT), and trust computing. She has contributed to interdisciplinary research, academic publications, and student mentoring in emerging computing technologies and innovative learning practices (E-mail: sornak@srmist.edu.in).

Dr. Muthukumar Manickam is an Assistant professor at the School of Computing, Department of Computer Science and Engineering, SRM Institute of Science and Technology, Tiruchirappalli. He holds a Ph.D. in Computer Science and Engineering from SRM Institute of Science and Technology, Kattankulathur, Chennai. With over 19 years of teaching experience, he has made notable academic contributions through the publication of 15 research papers in international journals and the presentation of 10 papers at conferences. His research interests primarily focus on computer networks, network security, and the broader cybersecurity landscape (E-mail: mkumar7680@gmail.com).

V. Preethi is an Assistant Professor in the Department of Networking and Communications, School of Computing, Faculty of Engineering and Technology at SRM Institute of Science and Technology (SRMIST). Her academic interests include data science, networking, and emerging computing technologies. She actively contributes to teaching, research, academic events, and student development initiatives in advanced computing domains (E-mail: preethiv3@srmist.edu.in).