IJIRIS: International Journal of Innovative Research in Information Security
Volume 12, Issue 07, October 2026
E-ISSN: 2349-7017
P-ISSN: 2349-7009
https://www.ijiris.com/archives
https://doi.org/10.26562/ijiris.2026.v1207.03

FedMedAI-ViT: Federated Learning with Homomorphic Encryption for Secure Multi-Disease Medical AI Diagnosis

Mrs.Jyothi M
Assistant Professor, Department of Information Science and Engineering
Vemana Institute of Technology, Bengaluru, India
jyothi.m@vemanait.edu.in https://ror.org/00ha14p11
https://orcid.org/0009-0007-9863-7913
Poojitha B H, Prerana G N, Shreya G Poojari, Shreya K
Department of Information Science and Engineering
Vemana Institute of Technology, Bengaluru, India
poojithabh.is2023@vemanait.edu.in, preranagn.is2023@vemanait.edu.in
shreyagpoojari.is2023@vemanait.edu.in, shreyak.is2023@vemanait.edu.in

Publication History
Manuscript Reference No: IJIRIS/RS/Vol.12/Issue07/OCIS10082
Research Article | Open Access | Double-Blind Peer-Reviewed| Article ID: IJIRAE/RS/Vol.12/Issue07/OCIS10082
Received:02,September 2026,Revised:28,September 2026,Accepted:03,October 2026,Published Online:30,October2026
https://www.ijiris.com/volumes/Vol12/iss-07/03.OCIS10082.pdf
Article Citation: Jyothi,Poojitha,Prerana,Shreya,Shreya(2026),FedMedAI-ViT: Federated Learning with Homomorphic Encryption for Secure Multi-Disease Medical AI Diagnosis. IJIRIS: International Journal of Innovative Research in Information Security, Volume 12, Issue 07 of 2026 pages 352-360
Doi:> https://doi.org/10.26562/ijiris.2026.v1207.03 **BibTeX Key: Jyothi@2026FedMedAI-ViT**
IJIRIS papers should be cited as IJIRIS(International Journal of Innovative Research in Information Security, AM Publications, India 2026, ISSN 2349-7017, https://doi.org/10.26562/ijiris.2026.v1207.03 The journal's official abbreviation is IJIRIS. **ORCiD:** https://orcid.org/0009-0004-9398-7488 About the License: Copyright©2026 copyright by the authors. This article is an open access and license under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/)

Abstract: Training diagnostic models across institutions without pooling patient images is attractive, but every privacy mechanism added to such a system needs its own evidence. This paper reports FedMedAI-ViT, a pilot federated prototype in which three simulated sites, holding brain MRI, EEG-spectrogram and retinal fundus images from public sources, train a shared patch-projection encoder with site-specific classification heads under the FedProx objective. The project was conceived around a pretrained ViT-B/16 backbone and encrypted aggregation of 768-dimensional embeddings. An audit of the code and logs shows that neither was used in the reported experiments, and the paper is written accordingly. A single-seed FedProx run reaches test accuracies of 80.39%, 86.54% and 45.60% on the three sites, a macro mean of 70.84% over 228 test images, with wide confidence intervals; the retinal result is close to the majority-class rate. Two further configurations are recorded but cannot be verified from the available outputs: Gaussian perturbation of client parameters, which carries no formal privacy guarantee, and a CKKS encrypt–decrypt round trip applied at inference. The paper separates implemented, intended, unverified and future components, and lists the experiments required before any privacy, efficiency or clinical claim can be made.

Index Terms: federated learning, FedProx, homomorphic encryption, CKKS, differential privacy, medical image classification, vision transformer, evidence audit

I. INTRODUCTION

Medical images are produced and held by individual institutions, and policies on patient data usually prevent hospitals from copying them to a central server. Federated learning (FL) lets sites train a shared model while the images stay local; only model parameters travel [1], [2]. Parameters are not harmless, however. Reconstruction attacks have recovered training images from shared gradients [3], [4], so FL on its own is not a privacy guarantee. Differential privacy (DP) [5] and homomorphic encryption (HE) [6] are the two mechanisms most often proposed to close that gap, and a recent preprint combines a Vision Transformer with CKKS encryption of the class-token embedding for histopathology [7]. The FedMedAI-ViT project set out to build a comparable multi-disease system: a shared Vision Transformer encoder, three disease-specific heads (brain tumor, epileptic seizure, diabetic retinopathy), FedProx aggregation [8], Gaussian noise, and CKKS encryption through TenSEAL [9], fronted by a Streamlit dashboard. This manuscript reports what the project actually did and what its recorded outputs support. The audit that preceded writing found that the reported experiments use a lightweight patch-projection encoder rather than a pretrained ViT-B/16, that the noise experiment perturbs model parameters rather than embeddings, and that CKKS was exercised as an inference-time round trip rather than inside federated aggregation. Those findings shape every section that follows. The central question is therefore modest: how does a shared encoder with site-specific heads behave under FedProx when three sites hold different modalities, and which of the proposed privacy components can currently be backed by evidence? The paper does not claim a pretrained or attention-based model, a formal privacy guarantee, measured resistance to gradient inversion, communication savings, clinical validity, or superiority over any published method.

E-ISSN: 2349-7017
P-ISSN: 2349-7009
https://www.ijiris.com/archives
https://doi.org/10.26562/ijiris.2026.v1207.03

Its contribution is a transparent pilot with verified per-site results, a clear statement of what each component does and does not show, and a list of the evidence still missing. The remainder of the paper follows the structure of the project. Sections II and III review related work and state the gap; Sections IV–VII describe the architecture, methods, implementation and evaluation protocol; Sections VIII–X cover the privacy analysis, experimental setup and results; and Sections XI–XIII give limitations, future work and conclusions.

II. RELATED WORK

A. Federated optimization and medical FL

McMahan *et al.* introduced federated averaging (FedAvg), in which clients train locally and a server averages their parameters, weighted by local sample counts [1]. Li *et al.* proposed FedProx, which adds a proximal term to each client's objective to limit drift from the current global model when client data or compute differ [8]. Rieke *et al.* discuss the promise and open problems of FL for digital health, including heterogeneity and privacy [2]. The Flower framework [10] provides the server–client machinery used in this project's integration tests.

B. Leakage from shared updates and differential privacy

Zhu *et al.* showed that private training inputs can be recovered from gradients by optimizing a dummy input to match them [3], and Geiping *et al.* extended this to higher-resolution images and trained networks [4]. These works motivate defenses but do not evaluate this project's model. Differential privacy offers a formal guarantee: a randomized mechanism is (ε, δ)-DP if output distributions on neighboring datasets differ by bounded factors [5]. In deep learning, DP-SGD obtains such a guarantee by clipping per-example gradients, adding calibrated noise, and tracking the cumulative privacy loss with an accountant [11]. Adding noise alone, without a bounded sensitivity and accounting, does not yield a stated (ε, δ).

C. CKKS and the base paper

CKKS supports approximate arithmetic on encrypted real vectors [6]; TenSEAL exposes it to Python tensor code [9]. Al Amin *et al.* report a federated ViT for lung-cancer histopathology across three clients in which the 768-dimensional class-token embedding is encrypted with CKKS before aggregation [7]. As reported by its authors, the framework reaches 96.12% global accuracy in the unencrypted domain and 90.02% in the encrypted domain, a gradient-reconstruction PSNR of 52.26 dB (SSIM 0.999) on unprotected gradients, a 30-fold communication reduction against gradient encryption, and 326 KB of encrypted transmission per round. That paper is this project's design inspiration. Its task, backbone, data and threat setting differ from ours, and its numbers are quoted only as the authors' reports; no comparison with the results below is made.

D. Vision Transformers

ViT-B/16 splits an image into 16 × 16 patches, embeds them linearly, adds position information and a class token, and processes the sequence with 12 transformer layers [12]. The encoder evaluated here keeps only a patch-embedding-like first stage and omits positional embeddings, class token and self-attention (Section IV-B).

III. RESEARCH GAP AND CONTRIBUTIONS

The project aimed at a multi-disease federation with layered privacy. The gap that matters for this paper is between that aim and the evidence in hand. Table I summarizes the foundations, the role each plays here, and the boundary of what it can support.

TABLE I- Literature foundations and their role in this project

Work	Established contribution	Role in this project	Boundary of applicability
FedAvg [1]	Parameter averaging across clients.	Aggregation rule (unweighted mean is used here).	Original form weights by client sample count; not the case here.
FedProx [8]	Proximal term limiting client drift.	Local objective in the reported run.	Convergence analysis does not cover this modality heterogeneous
Gradient leakage[3],[4]	Inputs can be reconstructed from gradients.	Motivation for privacy layers	No attack was run on this model.
DP and DP-SGD[5],[11]	Formal guarantee; clipping and accounting.	Reference point for what is *not* implemented.	Parameter noise here has no clipping or accounting.
CKKS, TenSEAL[6],[9]	Approximate arithmetic on encrypted vectors.	Inference-time encrypt/decrypt test.	Encrypted aggregation not integrated.
Al Amin *et al.* [7]	Reported ViT + CKKS class-token federation.	Design inspiration; results cited as reported.	Different task, backbone, data and threat model.
ViT [12]	Transformer over image patches.	Intended backbone (ViT-B/16).	Not used in any reported experiment.
Flower [10]	FL server/client framework.	Integration smoke tests.	No reported accuracy comes from it.

The contributions are deliberately limited:

1) A pilot federated prototype in which three sites with different modalities share one patch-projection encoder and keep site-specific heads, trained with a FedProx objective.
2) Verified per-site results from a single-seed run, reported with confidence intervals and majority-class reference rates, so their strength can be judged.
3) An explicit status for every proposed component implemented, intended, recorded but unverified, or future (Tables III and IV).

IJIRIS: International Journal of Innovative Research in Information Security
Volume 12, Issue 07, October 2026
E-ISSN: 2349-7017
P-ISSN: 2349-7009
https://www.ijiris.com/archives
https://doi.org/10.26562/ijiris.2026.v1207.03

4) A limited CKKS round-trip test and a description of the noise experiment as implemented, neither presented as a security result.
5) A prioritized list of the experiments needed to turn the pilot into evidence for the original aims.

IV. SYSTEM ARCHITECTURE

Fig. 1 shows the pipeline. Solid boxes were implemented and are part of the verified run; dashed amber boxes are experiment-specific additions whose results are recorded but not verifiable from the available outputs; the dotted grey band lists design intentions absent from all reported experiments.

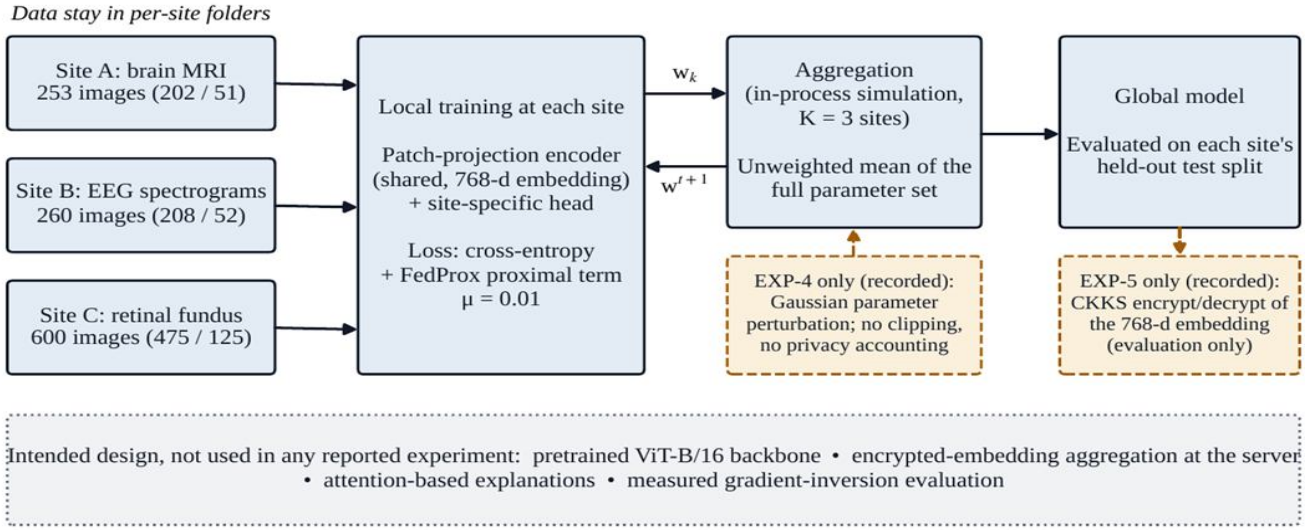


Fig.1. Pipeline as implemented and evaluated. The three "sites" are logical partitions of three public datasets simulated in one process; no real hospital participated. Dashed boxes are recorded but unverified configurations; the grey band lists intended, unimplemented components.

A. Sites and data flow

Each simulated site holds one modality: brain MRI (binary tumor/no tumor), EEG spectrograms (binary seizure/normal) and retinal fundus photographs (five severity grades). Raw images are read only from that site's folder during local training and testing. Because the sites differ in modality and label space rather than only in class balance, the federation is a multi-task one: only the encoder is common, and each head is trained by a single site.

B. Encoder and heads

The implemented encoder (Fig. 2) applies a 16 × 16, stride-16 convolution with 64 channels to a 224 × 224 × 3 input, flattens the 14 × 14 × 64 grid, projects it to 768 dimensions with a linear layer and ReLU, and applies layer normalization. Each head is a two-layer perceptron (768 → 256 → C, dropout 0.3). By arithmetic from this definition the model has about 10.28 M parameters, dominated by the 12 544 × 768 projection. The 768-dimensional output is called an embedding in this paper; it is not a transformer class token. The code also contains a path that would load the pretrained Hugging Face ViT-B/16, with an automatic fallback to this encoder if loading fails. That path was not exercised in any reported result, so ViT-B/16 is an intended backbone only.

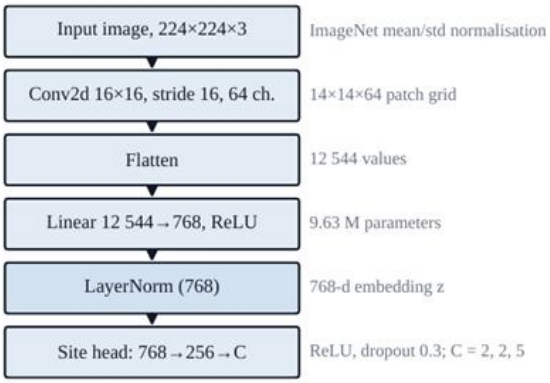


Fig.2. Implemented encoder and head. There is no positional embedding, class token or self-attention.

C. Federation

In the reported training script, federation is simulated in one process: each round every site copies the global parameters, trains locally, and a server step averages the three parameter sets. A separate Flower server and client implementation exists and was smoke-tested (Section VI), but the accuracy figures do not come from it.

D. Privacy components and their status

Two privacy-related components appear in the recorded experiments, both as additions to the verified pipeline. Gaussian perturbation is added to client parameters in the DP configuration; a CKKS encrypt–decrypt step is applied to embeddings at evaluation time in the final configuration.

IJIRIS: International Journal of Innovative Research in Information Security
Volume 12, Issue 07, October 2026
E-ISSN: 2349-7017
P-ISSN: 2349-7009
https://www.ijiris.com/archives
https://doi.org/10.26562/ijiris.2026.v1207.03

A helper that averages ciphertexts exists and is unit-tested on random vectors, but the aggregation step in the experiments never calls it. Encrypted aggregation is therefore intended, not implemented.

V. METHODOLOGY

A. Dataset construction and provenance

Table II lists the data. All three come from public repositories mirrored on GitHub because the original hosting platforms were unreachable from the build environment. The "hospital" names used in the project interface is labels for simulated sites.

TABLE II - Datasets sources and splits

Site	Task	Source and processing	Train / test
A	Brain MRI, 2 classes	253 images (155 tumor, 98 none); Kaggle data via a GitHub mirror [13].	202 / 51
B	EEG seizure, 2 classes	260 balanced rows of the UCI seizure data [14], [15]; 1-s signals converted to STFT spectrograms.	208 / 52
C	Retina, 5 grades	600 fundus photographs [16]; 10 source classes collapsed to 5 grades by the authors; resized to 400 px.	475 / 125

The retinal grade distribution is imbalanced (276, 99, 168, 34 and 23 images for grades 0–4 over the full set). Splits are random 80/20 partitions of image files; the EEG conversion script uses a fixed seed, while the procedure used to split sites A and C is not documented in the supplied files. No patient identifiers are available, so no patient-level separation could be enforced; Section XI discusses the consequence.

B. Preprocessing

Images are converted to RGB, resized to 224 × 224 and normalized with ImageNet channel means and standard deviations. No augmentation is applied. EEG signals are converted to spectrogram images by short-time Fourier transform before the split.

C. Local objective and aggregation

At round t site k starts from the global parameters w^t and minimizes

$$\mathcal{L}_k(w) = \mathbb{E}_{(x,y)\sim D_k}[\ell_{\mathrm{CE}}(h_k(f_\theta(x)), y)] + \frac{\mu}{2}\|w - w^t\|_2^2 \quad (1)$$

where f_θ is the shared encoder, h_k the site head, ℓ_{CE} cross-entropy, and the proximal norm runs over the full parameter set. For $\mu = 0$ this reduces to local training as used by FedAvg. The server then forms

$$w^{t+1} = \frac{1}{K}\sum_{k=1}^{K} w_k^{t+1}, \quad K = 3 \quad (2)$$

an unweighted mean of every parameter, including all three heads. Because site k's loss does not reach the other heads, each head receives its update averaged with two unchanged copies and moves by one third of its local step per round. This is a property of the implementation and may slow the heads' progress; its effect was not isolated experimentally.

D. Gaussian perturbation of parameters (DP configuration)

In the DP configuration, after local training each site adds independent Gaussian noise to every tensor whose name contains "weight" or "bias":

$$\tilde{w}_k = w_k + \eta, \quad \eta \sim \mathcal{N}(0, \sigma^2 I), \quad \sigma = \frac{10^{-3}}{\varepsilon_s}, \quad (3)$$

With $\varepsilon_s = 1$. The code calls ε_s "epsilon", but nothing bounds the sensitivity of the update, updates are not clipped, and no accountant is run, so ε_s is a noise-scale knob and not a privacy budget. No (ε, δ) guarantee follows from Eq. (3), and none is claimed. The project's documentation and result file describe this experiment as noise on clipped 768-dimensional embeddings with $\sigma = C/\varepsilon$; that mechanism exists in the code only on the dashboard inference path. This paper describes what the training code does.

E. CKKS round trip at inference

For the CKKS configuration, the 768-dimensional embedding z of each test image is encrypted and immediately decrypted before the site head is applied:

$$\hat{z} = \mathrm{Dec}_{sk}(\mathrm{Enc}_{pk}(z)), \quad \hat{y} = \arg\max h_k(\hat{z}), \quad (4)$$

using TenSEAL with polynomial modulus degree 8192, coefficient moduli of [60, 40, 40, 60] bits, global scale 2^{40} and Galois keys. One context object holds both public and secret keys in the same process, so the test shows numerical fidelity of the encoding, not a separation between a data holder and an untrusted server. The step is not part of training or aggregation.

VI. IMPLEMENTATION

The prototype is written in Python with PyTorch for modeling, Flower for the integration tests, TenSEAL for CKKS, scikit-learn for metrics, a Streamlit dashboard and a SQLite-backed FastAPI service for logging. Table III records what the supplied files show for each component and what is missing.

TABLE III - Implementation evidence in the supplied project files

Component	Status	Available evidence	Missing evidence
Patch-projection encoder, three heads	Implemented	Model code; per-site metric files and confusion matrices recomputed exactly.	Saved checkpoints; parameter count log.
FedProx local objective	Implemented	Proximal term in the training script and	Ablation isolating the term's effect

		in the Flower client.	
Training run (6 rounds)	Verified (single seed)	Per-round history ending at 70.84%; per-site metrics consistent with it.	Repeated seeds; raw prediction files.
Flower server/client	Smoke-tested	Logs: 2 rounds with 1 client; 3 clients completed round 1, then 2 of 3 failed in round 2 (reported as a CPU-time limit).	Complete multi-round run with accuracy.
Parameter perturbation	Recorded, unverified	Code applying Eq. (3); summary row.	Per-site outputs; repeated runs; accounting.
CKKS round trip	Recorded, unverified	Code and parameters; summary row; unit tests on random vectors.	Evaluation log; per-sample agreement; numerical error.
Encrypted aggregation	Not integrated	Helper functions for averaging ciphertexts.	Use in a training loop; key management.
Pretrained ViT-B/16	Intended	Loader with fallback to the lightweight encoder.	Any run that loads and trains it.
Attention-based explanation	Not claimed	Module exists; for this encoder it returns a placeholder.	Explanations from a real attention model.
Gradient-inversion demo	Illustrative only	Dashboard simulator producing synthetic reconstructions.	An actual attack on the trained model.
Dashboard and unit tests	Implemented	Streamlit app; 44 test functions defined in six files.	Test-run output.

VII. AI/ML EVALUATION METHODOLOGY

A. Classification metrics

For the binary sites, with the positive class being tumor or seizure, accuracy, precision, recall and F1 are

$$\mathrm{Acc} = \frac{TP + TN}{TP + TN + FP + FN}, \quad (5)$$

$$\mathrm{Prec} = \frac{TP}{TP + FP}, \quad \mathrm{Rec} = \frac{TP}{TP + FN}, \quad (6)$$

$$F_1 = \frac{2TP}{2TP + FP + FN}. \quad (7)$$

For the five-grade retinal task, precision, recall and F1 in the result files are support-weighted averages, and macro-averaged values are recomputed from the confusion matrix for this paper. The headline number is the macro mean of the three site accuracies, which gives each site equal weight regardless of test-set size. The pooled accuracy over all 228 test images is reported alongside.

$$A_{\mathrm{macro}} = \frac{1}{3} \sum_{k \in \{A,B,C\}} A_k, \quad (8)$$

B. Uncertainty and reference rates

With 51 to 125 test images per site, point estimates are coarse, so 95% Wilson score intervals are given for each accuracy:

$$\frac{\hat{p} + \frac{z^2}{2n} \pm z\sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{z^2}{4n^2}}}{1 + \frac{z^2}{n}}, \quad z = 1.96. \quad (9)$$

The majority-class rate of each test split, obtained from the confusion-matrix row sums, is shown as a reference that a trivial classifier would reach.

C. Conventions

All metrics are recomputed from the stored confusion matrices and compared with the stored values. AUC values for the binary sites are checked against the stored ROC points by trapezoidal integration. The retinal ROC curve in the result file is a fixed placeholder and is not used. There is one run with one seed, no validation split, and no statistical test between configurations.

VIII. SECURITY AND PRIVACY CONSIDERATIONS

A. THREAT MODEL

The motivating adversary is an honest-but-curious aggregator who observes the parameters each site sends and tries to reconstruct patient images [3], [4]. In both the in-process simulation and the Flower client, sites send plaintext parameters to the aggregator. Nothing implemented here prevents that adversary from reading them.

B. What the implemented components do and do not show

Parameter noise at $\sigma = 10^{-3}$ perturbs weights slightly, but without sensitivity bounding or accounting it supports no formal claim [11]. The CKKS test shows that decrypted embeddings remain usable by a head (if the recorded result holds); it does not show that the server never sees plaintext, because the secret key sits in the same process. The dashboard's attack simulator draws a scaled copy of the image with noise for the "unprotected" case and random pixels for the "protected" case, so its PSNR figures follow from construction and are not measurements. Table IV gives the resulting status of each claim.

IJIRIS: International Journal of Innovative Research in Information Security
Volume 12, Issue 07, October 2026
E-ISSN: 2349-7017
P-ISSN: 2349-7009
https://www.ijiris.com/archives
https://doi.org/10.26562/ijiris.2026.v1207.03

Table IV- Status of privacy and performance claims

Claim	Status
Raw images stay in site folders (simulation)	Implemented by design; not audited as a security property
FedProx proximal term applied	Implemented; verified run
Formal (ε, δ)-DP	Not supported
CKKS preserves predictions at inference	Recorded, unverified
Server aggregates ciphertexts	Not implemented
Resistance to gradient inversion	Not evaluated
Communication reduction	Not claimed
Attention-based explanations	Not claimed
Clinical validity or superiority	Not claimed

C. On the communication figure

The project files report a 323.39 KB ciphertext for one 768-dimensional vector and compare it with a constant of 9 794 KB labeled as the full update. That constant does not match the model (about 10.28 M parameters, roughly 40 000 KB in 32-bit floats), the two quantities describe different things, and the ciphertext figure comes from one random vector. No reduction factor is therefore claimed.

IX. EXPERIMENTAL SETUP

Table V lists the training configuration, and the five configurations were defined as follows. EXP-1 trains each site in isolation for 5 epochs. EXP-2 runs federated training with $\mu = 0$ for 5 rounds. EXP-3 uses $\mu = 0.01$ for 6 rounds. EXP-4 repeats FedProx for 5 rounds with the perturbation of Eq. (3). EXP-5 takes the EXP-4 model and evaluates it with the round trip of Eq. (4), without adding embedding noise. The numbers of rounds and epochs differ among configurations, so differences between them are confounded with training length. Model initialization is seeded with 42. The hardware used is not recorded in the supplied files, and training times are not reported.

TABLE V - Training configuration of the federated runs

Item	Value
Sites / clients	3 (one modality each)
Local epochs per round	2
Optimizer, learning rate	Adam, 10^{-4} (re-initialized each round)
Batch size	8
Proximal coefficient μ	0.01 (FedProx), 0 (FedAvg)
Aggregation	Unweighted mean, full parameter set
Rounds	6 (EXP-3), 5 (EXP-2, EXP-4)
Seed / repetitions	42 / one run
Validation split	None

X. RESULTS

A. Verified FedProx results

Table VI reports the FedProx run (EXP-3). Every value was recomputed from the stored confusion matrices and matches the stored metrics. Brain MRI reaches 80.39% (41 of 51) and seizure detection 86.54% (45 of 52); retinal grading reaches 45.60% (57 of 125). The macro mean is 70.84%, and the pooled accuracy over 228 images is 62.72% (95% interval 56.3–68.7). Fig. 3 shows the accuracies with their intervals and majority-class rates, and Fig. 4 shows the confusion matrices.

TABLE VI

Site (test *n*)	Acc. (%)	95% CI (%)	Maj. (%)	Prec.	Rec.	F1	Spec.	AUC
A: Brain MRI (51)	80.39	67.5–89.0	60.78	0.8182	0.8710	0.8438	0.7000	0.9065
B: Seizure (52)	86.54	74.7–93.3	53.85	0.8387	0.9286	0.8814	0.7917	0.9673
C: Retina (125)	45.60	37.1–54.3	44.80	0.3396	0.4560	0.3196	0.8072	0.6076
Macro mean	70.84	—	—	—	—	—	—	—

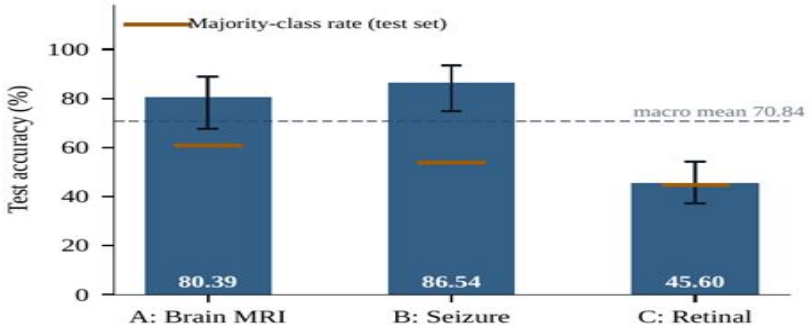


Fig.3. Test accuracy per site with 95% Wilson intervals (bars, whiskers) and majority-class rates of each test split (amber marks).

IJIRIS: International Journal of Innovative Research in Information Security
Volume 12, Issue 07, October 2026
E-ISSN: 2349-7017
P-ISSN: 2349-7009
https://www.ijiris.com/archives
https://doi.org/10.26562/ijiris.2026.v1207.03

Per-site results of the verified FedProx run (EXP-3, single seed). CI: 95% Wilson interval on accuracy; Maj.: majority-class rate; Prec.: precision; Rec.: recall; Spec.: specificity. For site C, precision, recall and F1 are support-weighted and specificity is the one-vs-rest macro mean; all were recomputed exactly from the confusion matrix. Binary AUCs were checked against the stored ROC points; site C's AUC (weighted one-vs-rest) is as stored and not recomputable

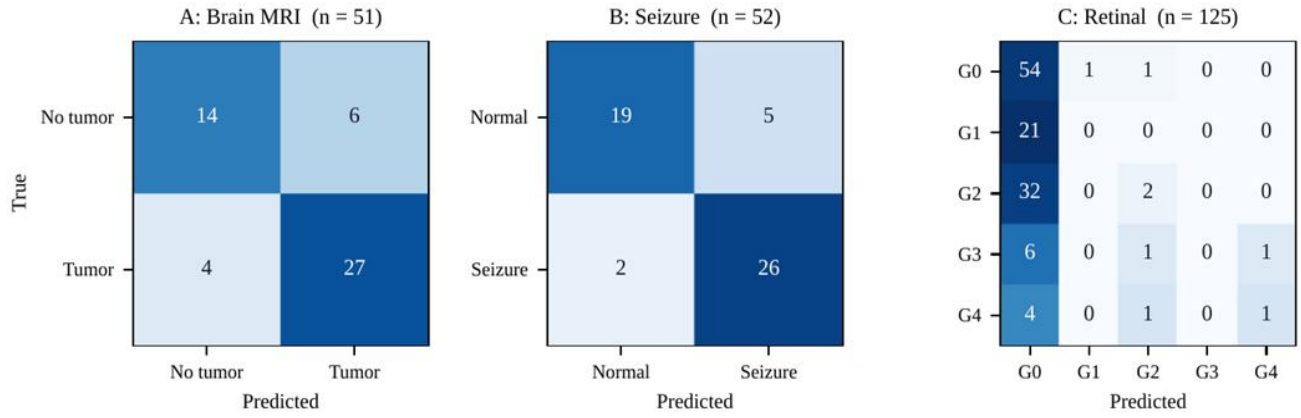


Fig. 4. Confusion matrices of the verified run (counts; shading is the row-normalized fraction). Retinal grades 1 and 3 are never predicted correctly, and almost all retinal images are assigned to grade 0. The binary sites exceed their majority-class rates by 20 to 33 points, which the intervals support. The retinal result is 0.8 points above the majority rate: of 125 images, 54 of 56 grade-0 images are correct, grades 1 and 3 have zero recall, and the macro-averaged F1 is 0.195 against a weighted F1 of 0.320. On this evidence the retinal head has learned little beyond the prior.

B. Round-by-round behavior

Fig. 5 shows the recorded history of the same run. Macro test accuracy is highest after round 1 (72.78%) and falls steadily to 70.84% after round 6 while the mean local training loss decreases from 0.79 to 0.56. The final round is the one reported; no round was selected. Possible reasons include the dilution of heads under Eq. (2), overfitting of the large projection layer on a few hundred images, and drift between modalities, but the available outputs cannot separate them.

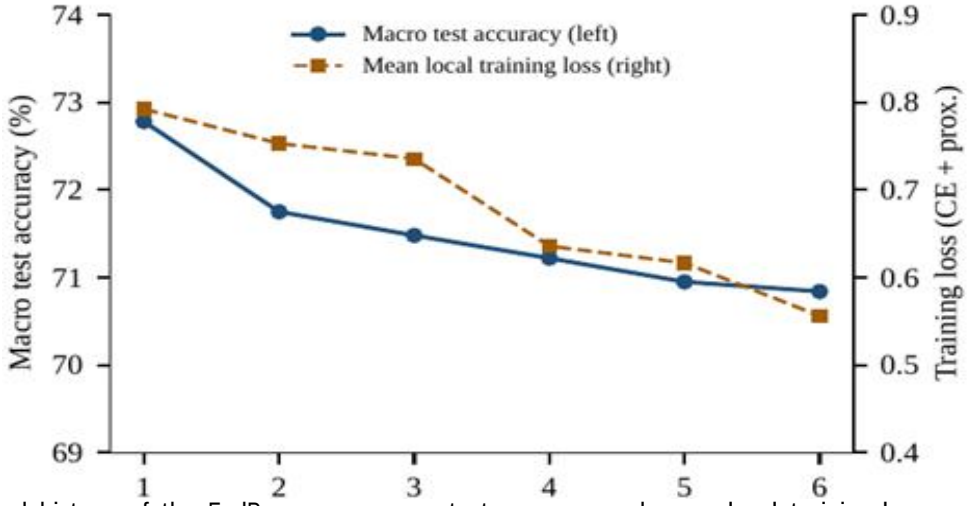


Fig.5. Recorded history of the FedProx run: macro test accuracy and mean local training loss per round. The final accuracy equals the verified 70.84%.

C. Recorded comparison of configurations

Table VII lists the macro accuracy recorded for all five configurations. Only EXP-3 can be traced to per-site outputs and a per-round history; the others are summary rows with no supporting files in the package and are reported as recorded.

Table VII - Recorded macro accuracy by configuration. Verification status refers to the files supplied for this study

ID	Configuration	Acc. (%)	Status
EXP-1	Isolated local training	74.05	Unverified
EXP-2	FedAvg (μ = 0)	72.40	Unverified
EXP-3	FedProx (μ = 0.01)	70.84	Verified
EXP-4	FedProx + parameter noise	71.32	Unverified
EXP-5	EXP-4 model + CKKS round trip	71.32	Unverified

Taking the recorded values at face value, isolated training scores highest and the federated and perturbed runs lie 1.7 to 3.2 points lower. With 228 test images, one image corresponds to about 0.4 points of pooled accuracy, the per-site intervals span 17 to 22 points, there is one seed, and training lengths differ, so none of these differences, including EXP-4 versus EXP-3, is distinguishable from noise. In particular, no benefit of FedProx over FedAvg, or of noise, can be inferred.

E-ISSN: 2349-7017
P-ISSN: 2349-7009
https://www.ijiris.com/archives
https://doi.org/10.26562/ijiris.2026.v1207.03

D. CKKS round trip

The recorded EXP-5 accuracy equals that of EXP-4, as expected if decrypted embeddings preserve predictions. The supplied files contain neither the evaluation output nor a measured decryption error; the unit tests only assert a mean absolute error below 0.01 on random vectors. The result is therefore stated as consistent with prediction preservation but unverified. Since EXP-5 adds no embedding noise, it also says nothing about the combined effect of noise and encryption. Analytically, unit-scale Gaussian noise added to a 768-dimensional embedding clipped to norm at most one, as in the dashboard path, has an expected norm near 27.7 against a signal norm of at most one, and the effect on accuracy of that path was not measured in the supplied outputs.

XI. LIMITATIONS AND THREATS TO VALIDITY

Model. The encoder is a patch projection with a linear layer, not a transformer; conclusions about ViT-B/16 or attention cannot be drawn. The parameter count of about 10.3 M on 202–475 training images per site invites overfitting.

Data. The datasets are small public mirrors, and splits are random over image files. The UCI seizure set consists of 23 one-second chunks from each of 500 individuals' recordings [14], and the brain MRI set may contain several slices per patient. Without patient identifiers, related images may fall on both sides of a split and inflate sites A and B. Site C's ten-to-five grade mapping was chosen by the authors and has not been checked by clinicians. The three sites are simulated partitions, not real institutions.

Evaluation. Results come from one seed with test sets of 51–125 images and no validation split. The configuration was developed over repeated runs observed against the same test splits, so the reported figures are development-set estimates, and some summary values moved between runs. Configurations differ in rounds and epochs.

Privacy. No formal guarantee, attack evaluation or key-management design exists, and parameters are exchanged in plaintext in both federation paths. The dashboard's privacy panel is illustrative and must not be read as measured protection.

Clinical. No clinician took part, and nothing here supports diagnostic use.

XII. FUTURE WORK

The first priority is a rerun with matched rounds and epochs, a validation split, patient- or recording-level partitions, and at least three seeds with mean and standard deviation. Next, the pretrained ViT-B/16 path should be trained under FedProx and compared with the present encoder and with a centralized baseline on the same data. A genuine DP experiment would clip per-site updates, calibrate noise to a stated (ε, δ) and track it with an accountant [11]. Encrypted aggregation should be integrated so that the server holds only ciphertext and only data holders hold secret keys, with measured accuracy, latency and payload against the true size of the transmitted update. A real gradient-inversion attack on the trained model, with and without each defense, would replace the illustrative simulator. Larger datasets, sample-weighted aggregation and per-site head handling that avoids dilution should follow, and any attention-based explanation must be validated against a model that actually uses attention.

XIII. CONCLUSION

FedMedAI-ViT is reported here as a pilot federated prototype whose evidence is narrower than its original ambition. A shared patch-projection encoder with site-specific heads, trained with a FedProx objective on three public datasets, reaches 80.39%, 86.54% and 45.60% on the three test splits (macro mean 70.84%), with intervals wide enough that configuration differences cannot be resolved and with the retinal task close to the majority rate. The parameter-noise and CKKS experiments are recorded but unverified, the former carries no formal privacy guarantee, and the latter is an inference-time round trip outside training and aggregation. Pretrained ViT-B/16, encrypted aggregation, measured attack resistance and communication savings remain intended work. The prototype is a starting point, and the list of missing evidence defines what must be done before the privacy-preserving, multi-disease claims in the project's title can be supported.

ACKNOWLEDGMENT

The authors thank Mrs. Jyothi M, Assistant Professor, Department of Information Science and Engineering, Vemana Institute of Technology, for her guidance and support during the project. The authors also acknowledge the department and institution for supporting the project work.

AUTHOR CONTRIBUTION STATEMENT

Conceptualization: Jyothi M
Literature Review and Methodology design: Poojitha B H, Prerana G N, Shreya G Poojari, Shreya K
Software: Poojitha B H, Prerana G N, Shreya G Poojari, Shreya K
Validation: Poojitha B H, Prerana G N, Shreya G Poojari, Shreya K
Formal Analysis: Poojitha B H, Prerana G N, Shreya G Poojari, Shreya K
Investigation: Jyothi M
Resources: Poojitha B H, Prerana G N, Shreya G Poojari, Shreya K
Data Curation: Poojitha B H, Prerana G N, Shreya G Poojari, Shreya K
Writing original draft preparation: Poojitha B H, Prerana G N, Shreya G Poojari, Shreya K
Writing review and Editing: Jyothi M
Visualization: Jyothi M
Supervision: Jyothi M

Conflict of interest: The authors declare no conflicts of interest.
Data availability statement: Data supporting these findings are available within the article, at https://doi.org/10.26562/ijiris.2026.v1207.03, or upon request.
Supplementary materials: The supplementary materials are available at https://doi.org/10.26562/ijiris.2026.v1207.03
Publisher's note: AM Publications, India. IJIRIS (International Journal of Innovative Research in Information Security) Journals stays neutral with regard to jurisdictional claims in published maps and institutional affiliations. All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.
https://ampublications-india.com/

REFERENCES

1. Madhu,Anand,Aditya,Sara,Rakshith(2024).Predictive Analysis and Discerning of Polycystic Ovary Syndrome in Women using Machine Learning Computations. International Journal of Innovative Research in Information Security, Volume 10, Issue 04, Pages 178-184 https://doi.org/10.26562/ijiris.2024.v1004.11
2. N.Rieke, J.Hancox, W.Li, F.Milletarì, H.R.Roth, S.Albarqouni, S.Bakas, M.N.Galtier, B.A.Landman, K.Maier-Hein, S. Ourselin, M.Sheller, R.M. Summers, A.Trask, D.Xu, M.Baust, and M.J.Cardoso, "The future of digital health with federated learning," npj Digital Medicine, vol. 3, p. 119, 2020. https://doi.org/10.1038/s41746-020-00323-1
3. L.Zhu, Z.Liu, and S.Han, "Deep leakage from gradients," in Advances in Neural Information Processing Systems (NeurIPS), 2019, pp. 14774–14784.
4. Pavan,Pruthvi,Salma,Shreya,Sujana(2024). Revolutionizing Deepfake Detection and Realtime Video Vision with CNN-based Deep Learning Model. International Journal of Innovative Research in Information Security, Volume 10, Issue 03, Pages 173-177 https://doi.org/10.26562/ijiris.2024.v1004.10
5. C.Dwork and A.Roth, "The algorithmic foundations of differential privacy," Found. Trends Theor. Comput. Sci., vol. 9, no. 3–4, pp. 211–407, 2014. https://doi.org/10.1561/0400000042
6. J.H.Cheon, A.Kim, M.Kim, and Y.Song, "Homomorphic encryption for arithmetic of approximate numbers," in Advances in Cryptology – ASIACRYPT 2017, 2017, pp. 409–437. https://doi.org/10.1007/978-3-319-70694-8_15
7. Al Amin, K.Hasan, L.Hong, and S.Ullah, "Privacy-preserving federated vision transformer learning leveraging lightweight homomorphic encryption in medical AI," arXiv: 2511.20983 [preprint], Nov. 2025. https://doi.org/10.1109/ICNC68183.2026.11416875
8. Divya,Kalluri,Krishna,Kancham,Hari(2024). Computer Vision-Based Mobile Application for Visually Impaired. International Journal of Innovative Research in Information Security, Volume 10, Issue 03, Pages 156-161 https://doi.org/10.26562/ijiris.2024.v1004.07
9. Rashmi,Kavya,Deepthi,Konuri(2024). Load Balancing using Genetic Algorithm with Cloud SIM Simulator in Open Source. International Journal of Innovative Research in Information Security, Volume 10, Issue 03, Pages 162-167 https://doi.org/10.26562/ijiris.2024.v1004.08
10. Shilpa,Aravind,Dilip,Hinabinhu(2024). Transforming the University HR on Boarding Process with RPA. International Journal of Innovative Research in Information Security, Volume 10, Issue 03, Pages 168-172 https://doi.org/10.26562/ijiris.2024.v1004.09
11. M.Abadi, A.Chu, I.Goodfellow, H.B.McMahan, I.Mironov, K.Talwar, and L. Zhang, "Deep learning with differential privacy," in Proc. ACM SIGSAC Conf. Computer and Communications Security (CCS), 2016. https://doi.org/10.1145/2976749.2978318
12. A.Dosovitskiy, L.Beyer, A.Kolesnikov, D.Weissenborn, X.Zhai, T.Unterthiner, M.Dehghani, M.Minderer, G.Heigold, S. Gelly, J.Uszkoreit, and N.Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale," arXiv: 2010.11929 [preprint], 2020.
13. M.A.Habib, "Brain-tumor-detection (repository; mirrors the kaggle "brain mri images for brain tumor detection" data)," GitHub, https://github.com/MohamedAliHabib/Brain-Tumor-Detection, accessed: Oct. 14, 2026.
14. Q.Wu,E.Fokoue, "Epileptic seizure recognition," UCI Machine Learning Repository, https://archive.ics.uci.edu/dataset/388/epileptic+seizure+recognition, accessed: Oct. 14, 2026.
15. R.G.Andrzejak, K.Lehnertz, F.Mormann, C.Rieke, P.David, and C.E.Elger, "Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state," Phys. Rev. E, vol. 64, no. 6, p. 061907, 2001. https://doi.org/10.1103/PhysRevE.64.061907
16. S.S.Rahim,"Retinal fundus image dataset for diabetic retinopathy detection," GitHub, https://github.com/SRahimDR/DR-image-dataset, 2025, accessed: Oct. 14, 2026.