International Journal on Advanced Computer Engineering and Communication Technology

ISSN: 2278-5140

Volume 15 Issue 2 (2026) | Pages: 316-324

Trustworthy Federated Clinical Intelligence with Homomorphic Encryption and Explainable AI

Abhay R. Gaidhani[1], Anmol S. Budhewar[2], Pranav Nitin Pagare[3], Ekansh Udaykumar Tayade[4], Om Sunil Kute[5], Darshan Nandkishor Aher[6]

[1,2,3,4,5,6]Department of Computer Engineering, Sandip Institute of Technology and Research Centre, Nashik, Maharashtra, India
[1]*abhay.gaidhani@sitrc.org*, [2]*anmolsbudhewar@gmail.com*, [3]*pranavnpagare@gmail.com*, [4]*ekanshshweta07@gmail.com*, [5]*omkute04@gmail.com*, [6]*aherdarshan2803@gmail.com*

Peer Review Information

***Type**: Article*
***Received**: 04 Aug 2026*
***Revised**: 03 Sept 2026*
***Accepted**: 09 Oct 2026*
***Published**: 02 Nov 2026*

Abstract

Cross-institutional clinical artificial intelligence is constrained by fragmented patient data, non-identically distributed (non-IID) populations, leakage from shared model updates, and limited interpretability. Federated learning (FL) preserves data locality, but ordinary FL neither makes model updates confidential nor guarantees stable optimization under heterogeneous clients. This paper proposes a defense-in-depth framework that coordinates Federated Proximal (FedProx) optimization, record-level differential privacy (DP), Cheon–Kim–Kim–Song (CKKS) approximate homomorphic encryption, and explainable artificial intelligence (XAI). FedProx limits local model drift; DP-SGD bounds the influence of an individual training record; and CKKS permits weighted aggregation while client updates remain encrypted. A key-separated design keeps the aggregation server from being provisioned with the secret key used to reveal individual contributions; only the aggregate is released for global-model construction. Grad-CAM and SHAP/LIME provide clinician-facing explanations, while an audit dashboard reports convergence, privacy-accounting inputs, encryption status, and systems overhead. We formalize the learning objective, privacy mechanism, encrypted aggregation, threat model, numerical approximation error, communication footprint, and an evaluation protocol spanning IID and controlled non-IID partitions. The contribution is a reproducible system-level co-design and evaluation framework rather than a new cryptographic primitive or optimization algorithm.

Keywords: Federated Learning; CKKS; Homomorphic Encryption; Differential Privacy; FedProx; Explainable AI; Clinical Decision Support; Non-IID Data

How to Cite This Article

Gaidhani, A. R., Budhewar, A. S., Pagare, P. N., Tayade, E. U., Kute, O. S., & Aher, D. N. (2026). Trustworthy federated clinical intelligence with homomorphic encryption and explainable AI. *International Journal on Advanced Computer Engineering and Communication Technology,* 15(2), 316–324.

2278-5140/ © 2026 The Authors. This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0).

Introduction

Clinical artificial intelligence increasingly relies on distributed imaging and physiological data, yet transferring patient records to a centralized repository raises privacy, governance, and security concerns [1], [2], [6], [29]. FL addresses the data-locality problem by repeatedly distributing a model, optimizing it at participating institutions, and aggregating model information instead of raw records [3], [4]. This architecture is attractive for cross-silo healthcare, but data locality alone is not a complete privacy or robustness guarantee.

Three limitations are especially important. First, hospital datasets are naturally non-IID because patient populations, prevalence, acquisition devices, and annotation practices differ. FedAvg can therefore experience client drift; FedProx constrains local optimization around the current global model and improves stability under systems and statistical heterogeneity [5]. Second, gradients and model parameters can disclose training information even when raw records never leave a site [19]. DP provides a formal mechanism for limiting record influence [17], [18], whereas encrypted aggregation protects the numerical update itself from an aggregation service. Third, black-box predictions are difficult to audit in clinical workflows, motivating local explanation methods such as Grad-CAM, LIME, and SHAP [20]–[22].

The framework of Qazi et al. combines federated optimization, differential privacy, secure communication, and XAI across multiple medical-imaging tasks [1]. Our work follows that system-level direction but changes the confidentiality boundary: rather than protecting an update only while it is transported, the proposed protocol uses CKKS approximate homomorphic encryption so that a weighted sum can be evaluated before update decryption. CKKS supports packed approximate arithmetic on numerical vectors [14], and encrypted cross-silo FL systems such as BatchCrypt demonstrate why encryption cost and ciphertext expansion must be measured explicitly [15].

The contributions are fourfold: (1) a unified FedProx–DP–CKKS–XAI clinical FL architecture; (2) a key-separated encrypted-aggregation protocol that clearly distinguishes update confidentiality from record-level DP; (3) an explicit threat, numerical-error, and communication analysis; and (4) a reproducible evaluation matrix covering IID, label-skew, quantity-skew, and feature-skew conditions. No unmeasured accuracy or privacy claim is assumed in this framework paper.

Related Work and Research Gap

FL has been studied for multi-institutional healthcare, retinal disease detection, diabetic-retinopathy analysis, brain-tumor classification, and collaborative clinical prediction [7]–[11], [13]. Surveys identify privacy, heterogeneity, communication overhead, and deployment governance as persistent concerns [3], [6], [12]. Personalized approaches for multi-center tuberculosis diagnosis further illustrate the value of explicit adaptation when institutional distributions differ [26], [27].

Recent work by Budhewar and collaborators provides two locally relevant system perspectives. Their neighbour-aware semi-supervised decentralized learning study explicitly considers device-to-device cooperation, model quality, data-distribution similarity, and communication overhead in collaborative learning [31]. In healthcare, their Secure Care Hub architecture uses decentralized security mechanisms to strengthen patient-record integrity, privacy, and controlled access [32]. More recently, Budhewar et al. examined federated AI-driven digital twins for healthcare, with privacy and clinical intelligence treated as architectural concerns rather than afterthoughts [33]. These studies reinforce the need to evaluate distributed learning, confidentiality, and operational cost jointly. In medical imaging, class imbalance and acquisition variability remain additional concerns in tuberculosis and retinal-disease analysis [23]–[25], [28].

DP and encrypted aggregation solve different problems. DP-SGD clips per-example gradients and injects calibrated noise to bound record influence [18]; secure aggregation seeks to reveal aggregate information rather than individual client contributions [16]. CKKS provides an encrypted-computation layer for approximate numerical vectors [14]. XAI is orthogonal to both: Grad-CAM localizes influential image regions, whereas LIME and SHAP explain local predictions through surrogate or additive attribution models [20]–[22].

Table 1 motivates the gap we address here: existing systems optimise for a subset of the heterogeneity, privacy, encrypted computation, interpretability, and operational auditability. Our novelty is not the reinvention of their individual algorithms but the coordinated design and joint evaluation of these requirements.

Table 1. *System-Level Comparison of Representative Approaches*

Method	Non-IID	DP	Enc. Agg.	XAI	Audit
FedAvg [4]	–	–	–	–	–
FedProx [5]	✓	–	–	–	–
Gulati et al. [8]	✓	–	–	–	–
BatchCrypt [15]	–	–	✓	–	–

Qazi et al. [1]	✓	✓	–	✓	–
Proposed	✓	✓	✓	✓	✓

Proposed Privacy-Preserving Framework

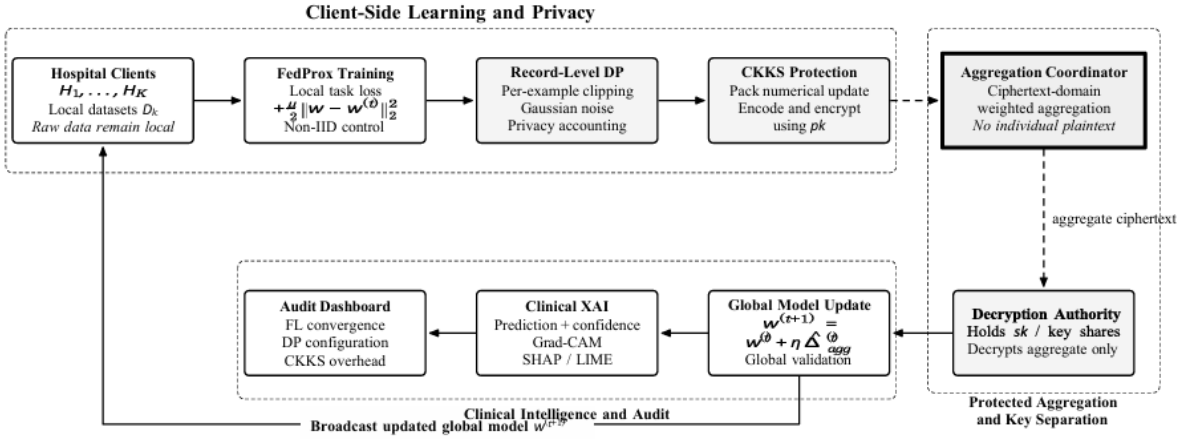


Fig. 1. *Defense-in-depth reference architecture for trustworthy federated clinical intelligence.*

Participating hospitals retain raw clinical data locally and use FedProx to reduce non-IID client drift. Record-level differential privacy limits individual-record influence, while CKKS protects numerical model updates during aggregation. A separate decryption authority releases only the aggregate update. The resulting global model supports explainable clinical decision support and auditable system monitoring.

Architecture and Trust Boundaries

Let $D_k = \{(x_i, y_i)\}_{i=1}^{n_k}$ be the datasets for K hospitals. Raw images, signals and labels do not leave the hospital environment. The coordinator is modelled as honest-but-curious, It follows the protocol but it may try to inspect the updates that it receives. In the baseline model, clients are protocol-compliant. Malicious-client poisoning, compromised endpoints, denial-of-service, collusion with the decryption authority, etc. are outside the primary threat model and are reported as residual risks and not silently excluded.

The participating clients and the coordinator have access to a common CKKS public key, pk. The secret key sk is owned by a logically separate decryption authority, or, in a stronger deployment, shared thru threshold key shares. The coordinator submits only the aggregate ciphertext for decryption. This separation is essential because homomorphic encryption does not protect individual updates from a party that also possesses the decryption key.

FedProx Optimization

For model parameters $w \in \mathbb{R}^P$ and task loss ℓ, client k has empirical risk

$$F_k(w) = \frac{1}{n_k}\sum_{(x_i,y_i)\in D_k} \ell\,(f_w(x_i), y_i). \qquad (1)$$

At round t, FedProx solves

$$w_k^{(t+1)} = \arg\min_w \left[F_k(w) + \frac{\mu}{2}\left\|w - w^{(t)}\right\|_2^2\right], \qquad (2)$$

where $\mu \geq 0$ controls the local deviation from the broadcast model $w^{(t)}$ [5]. The proximal term addresses optimization heterogeneity; it is not a confidentiality mechanism.

Record-Level Differential Privacy

For a local mini-batch B, let g_i be the per-example gradient of the data-dependent loss. DP-SGD clips each gradient before adding Gaussian noise:

$$\bar{g}_i = \frac{g_i}{\max(1, \|g_i\|_2 / C)} \qquad (3)$$

$$\tilde{g} = \frac{1}{|B|}\left[\sum_{i\in B} \bar{g}_i + \mathcal{N}(0, \sigma^2 C^2 I)\right]$$

The non-data-dependent FedProx regularizer is then included in the local optimization step. A randomized mechanism M is (ϵ, δ)-DP when

$$\Pr[M(D) \in S] \leq e^{\epsilon} \Pr[M(D') \in S] + \delta \qquad (4)$$

for neighboring record-level datasets D, D' and measurable S [17]. Because local records are accessed repeatedly, the final privacy loss must be computed with an explicit accountant using sampling rate, noise multiplier \(\sigma\), and the number of optimization steps [18], [34]. If an implementation clips only a final client update, its privacy unit and sensitivity model differ; such a mechanism must not be reported as record-level DP without a corresponding proof/accountant.

Once a local release satisfies DP under stated assumptions, CKKS encoding, encryption, and subsequent aggregation are post-processing with respect to that release and do not by themselves weaken the established DP guarantee. Repeated training rounds, however, still require composition accounting.

CKKS Homomorphic Aggregation

After private local optimization, client k forms $\Delta_k^{(t)} = w_k^{(t+1)} - w^{(t)}$. Let s be the usable CKKS slot capacity and \(m=\lceil P/s\rceil\) the number of ciphertexts required for a P-parameter update. Client k computes

$$c_k^{(t)} = \mathrm{Enc}_{pk}\big(\mathrm{Encode}\big(\Delta_k^{(t)}\big)\big). \qquad (5)$$

For selected clients S_t, the coordinator evaluates

$$c_{\mathrm{agg}}^{(t)} = \sum_{k \in S_t} \alpha_k \odot c_k^{(t)}, \qquad \alpha_k = \frac{n_k}{\sum_{j \in S_t} n_j}, \quad (6)$$

where $\odot$ denotes multiplication by a plaintext scalar represented consistently with the CKKS scale. The key authority releases only

$$\widehat{\Delta}_{\mathrm{agg}}^{(t)} = \mathrm{Decode}\big(\mathrm{Dec}_{sk}\big(c_{\mathrm{agg}}^{(t)}\big)\big), \quad w^{(t+1)} = w^{(t)} + \eta \widehat{\Delta}_{\mathrm{agg}}^{(t)}. \qquad (7)$$

Unlike ordinary transport encryption, Eq. (6) is evaluated while individual updates remain encrypted. Because the protocol needs additions and plaintext-scalar products rather than deep encrypted inference, its multiplicative depth is shallow; nevertheless, CKKS polynomial degree, modulus chain, scale, and packing strategy must be selected with a security-compliant library profile and then reported exactly. They should not be hard-coded before the model dimension and target precision are known.

CKKS is approximate. Numerical fidelity is therefore a measured quantity:

$$E_{\mathrm{CKKS}} = \frac{\|\Delta_{\mathrm{plain}} - \widehat{\Delta}_{\mathrm{CKKS}}\|_2}{\|\Delta_{\mathrm{plain}}\|_2 + \tau}, \quad (8)$$

where $\tau > 0$ avoids in stability for near-zero refe rence vectors. This error is separated from the stochastic distortion intentionally introduced by DP.

Table 2. *Security And Optimization Roles of Framework Components*

Layer	Primary property	Does not by itself solve
Data locality	raw records stay on site	gradient/update leakage
FedProx	non-IID drift control	confidentiality or DP
DP-SGD	bounded record influence	ciphertext confidentiality
CKKS	update confidentiality during aggregation	poisoning or memorization
XAI/audit	interpretability and traceability	cryptographic protection

Explainability and Audit

For convolutional feature map A^k and target class score y^c, Grad-CAM computes

$$\alpha_k^c = \frac{1}{Z} \sum_{i,j} \frac{\partial y^c}{\partial A_{ij}^k}, \qquad L_{\mathrm{GradCAM}}^c = \mathrm{ReLU}\Big(\sum_k \alpha_k^c A^k\Big) \quad (9)$$

to obtain a class-specific saliency map [20]. SHAP or LIME can provide complementary local attributions where computationally feasible [21], [22]. Explanations are treated as audit evidence rather than proof of clinical causality.

The dashboard separates a clinical view (input preview, predicted class, calibrated confidence, and explanation) from a research-administration view (round number, participating clients, local/global metrics, clipping norm, noise multiplier, privacy-accounting result, CKKS parameter summary, serialized ciphertext size, encryption/decryption time, and audit logs). This separation avoids exposing internal security configuration to ordinary clinical users.

Protocol Summary

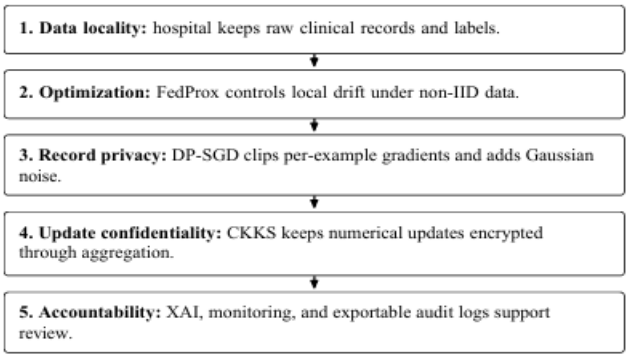


Fig. 2. *Defense-in-depth interpretation.*

The layers are complementary; no individual component is claimed to provide end-to-end security.

Algorithm 1: DP–CKKS FedProx Training

Algorithm 1: DP–CKKS FedProx Training

Initialize $w^{(0)}$; distribute pk; protect sk
for $t = 0, \ldots, T-1$ **do**
  Select S_t and broadcast $w^{(t)}$
  for all $k \in S_t$ **in parallel do**
    Optimize Eq. (2) using DP-SGD in Eq. (3)
    Form $\Delta_k^{(t)}$, CKKS-pack, encode, and encrypt
    Send ciphertext vector $c_k^{(t)}$ to coordinator
  end for
  Evaluate encrypted weighted sum in Eq. (6)
  Decrypt aggregate only; update via Eq. (7)
  Record utility, privacy, CKKS, and communication metrics
end for
Generate hospital-side XAI and export the audit report

Fig. 3. *One end-to-end training protocol.*

Evaluation Methodology

Dataset and Client Partitioning

The architecture is modality-agnostic and can be evaluated on the tasks used by the base study—tuberculosis chest X-rays, diabetic-retinopathy fundus images, or brain-tumor MRI [1]. The implementation should select a verified public dataset, document its license and exact class counts, remove corrupted or duplicate samples where feasible, and create fixed train/validation/test partitions. Patient-independent splitting is required whenever patient identifiers are available because image-level random splitting can otherwise leak subject-specific information across sets.

Four client-distribution regimes are proposed. IID uses stratified allocation with similar class proportions. Label skew samples client class proportions from

$$\pi_k \sim \text{Dirichlet}(\alpha \mathbf{1}_C), \qquad (10)$$

where smaller α produces stronger imbalance. Quantity skew varies n_k across hospitals. Feature skew applies controlled acquisition-style transforms such as contrast or noise changes while preserving labels. The exact random seed and partition indices are saved so that all baselines see identical data.

Medical tasks should not be summarized by accuracy alone. TB and retinal studies demonstrate that class imbalance, lesion subtlety, and domain variation can materially affect interpretation of performance [23]–[25], [28]. Accordingly, macro-F1, balanced accuracy, sensitivity, specificity, and AUROC are preferred, with per-class reporting where applicable.

B. Baselines, Ablations, and Fairness Controls

The ablation sequence is

$$Centralized \rightarrow FedAvg \rightarrow FedProx \rightarrow FedProx + DP \rightarrow FedProx + CKKS \rightarrow Full\ Framework. \qquad (11)$$

All variants use the same backbone, data split, augmentation policy, optimizer family, local epochs, and validation criterion wherever the comparison permits. Hyperparameters such as μ, learning rate, and early stopping are selected using validation data only. DP variants additionally disclose C, σ, sampling model, accountant, and final (ϵ, δ). CKKS variants disclose polynomial modulus degree, coefficient-modulus chain, scale, packing, serialized bytes, and measured error from Eq. (8).

To isolate the effect of non-IID learning, each ablation is repeated under at least IID and one controlled label-skew setting. Where compute allows, quantity and feature skew are added as separate experiments rather than mixed into a single ambiguous "non-IID" condition. Results should report mean and dispersion across repeated random seeds rather than a single favorable run.

The experiments are organised around five research questions. RQ1: Is the convergence and macro-level utility of FedProx better than FedAvg with the increasing heterogeneity? RQ2 measures the impact of tightening the privacy budget on record-level DP, sensitivity, macro-F1 and calibration. RQ3 explores the higher latency, memory and communication costs of CKKS and confirms that Eq. (8) is negligible relative to the stochasticity of training. RQ4 verifies the target trust boundary by demonstrating that the coordinator can compute Eq. (6) without knowledge of sk or receiving client updates in the clear. RQ5 studies the qualitative stability of the representative explanations for simple and privacy-preserving models, including true-positive, true-negative, false-positive and false-negative cases.

For utility comparisons, each reported metric should include the number of independent runs and either a standard deviation or a non-parametric confidence interval. Pairwise claims should be supported by a test appropriate to the repeated experimental design rather than inferred from overlapping point estimates. This is particularly important under non-IID partitioning because a different client seed can materially alter class exposure at individual hospitals.

Communication and Computational Cost

Let the CKKS ring degree be N, the active coefficient modulus contains Q bits in total, and m ciphertexts represent one packed update. A first-order estimate for a two-polynomial CKKS ciphertext payload is

$$B_{\mathrm{ct}} \approx \frac{2mNQ}{8} \text{ bytes}, \qquad (12)$$

excluding headers, compression, and library-specific serialization. The experiment must therefore report measured serialized bytes rather than relying only on Eq. (12). For a client with n_k samples, batch size B, E local epochs, and per-batch model cost C_f, local training is approximately $O\left(E\lceil n_k/B\rceil C_f\right)$. DP-SGD adds per-example clipping/noise overhead, while CKKS adds packing, encryption, network transfer, ciphertext arithmetic, and aggregate decryption. Wall-clock latency and peak memory are consequently first-class measurements [12], [15].

Resource efficiency also matters in decentralized clinical learning. Budhewar et al. explicitly balance neighbour quality and communication overhead in semi-supervised decentralized machine learning [31]. We use that systems perspective only to motivate explicit reporting of communication, latency, and memory costs; their neighbour-selection method is not claimed as a component of the proposed FL protocol.

Table 3. *Required Evaluation Dimensions*

Axis	Report	Purpose
Utility	macro-F1, AUROC, balanced accuracy, sensitivity	diagnostic performance
Convergence	loss, rounds-to-target, client divergence	non-IID stability
Privacy	C, σ, ϵ, δ, accountant	privacy–utility trade-off
CKKS	enc./dec. time, E CKKS, expansion	protected-compute cost
Systems	bytes/round, latency, peak memory	cross-silo feasibility
XAI	representative cases, stability/sanity checks	error interpretation quality

Implementation Profile

The framework can be implemented with a standard deep-learning stack, an FL orchestrator, a DP library that exposes per-example clipping and accounting, and a CKKS library such as Microsoft SEAL/TenSEAL or OpenFHE. The cryptographic layer is deliberately limited to update aggregation: no encrypted neural-network inference is required. Consequently, the aggregation primitive uses ciphertext additions and plaintext-scalar products and does not require ciphertext–ciphertext multiplication or bootstrapping. Scale alignment and rescaling are nevertheless implementation-dependent and must be validated with a plaintext-versus-CKKS aggregation test before long FL runs.

A practical profiling sequence is: first train the centralized and plain-FL baselines; next tune FedProx under the selected non-IID partition; then add DP and establish the privacy–utility curve; finally enable CKKS and profile packing, serialized size, encryption/decryption latency, and Eq. (8). This order prevents a cryptographic bottleneck or an over-aggressive privacy setting from being misdiagnosed as a learning failure. The implementation should also avoid encrypting optimizer state or layers that are never transmitted; only the update representation required by the chosen FL protocol is packed.

Reproducibility and Reporting

A reproducible run record includes dataset version, partition seed, model hash, optimizer and scheduler, local epochs, selected-client fraction, number of rounds, FedProx μ, DP accountant inputs and result, CKKS configuration, hardware, software/library versions, and measured timing. Audit artefacts should include only non-sensitive aggregate metrics and configuration metadata. Figures should use clinical example images from public test data, and be identified as illustrative evidence rather than diagnostic evidence.

Security, Privacy, And Clinical Discussion

The framework provides a systematic separation of assurances Data locality reduces direct exposure, but not update leakage. FedProx improves optimisation, but at the cost of privacy. DP bounds the contribution of a record, but does not produce an update ciphertext-confidential. CKKS hides numeric updates from an uncoordinated party that does not have sk, but does not protect against memorisation, model poisoning, or endpoint compromise. XAI and logging are good for accountability but explanation artefacts can also be sensitive themselves when derived from private examples. Hence, Table II characterises each layer in terms of its main property and boundary.

The key-separated architecture reduces the risk of individual-update inspection by the coordinator. It does not, however, eliminate all leakage from an aggregate, especially for very small participating sets or colluding parties. A stronger implementation can combine minimum cohort-size policies, authenticated clients, threshold decryption, signed model releases, and dedicated secure-aggregation protocols [16]. These controls should be selected according to the deployment threat model rather than added as unsupported security labels.

There are also clinical and statistical limitations. Stronger DP typically increases stochastic distortion and can disproportionately affect minority-class sensitivity. Severe non-IID data may remain difficult even with FedProx. CKKS increases memory, communication, and latency and introduces bounded approximation error. Saliency or attribution maps can appear plausible without being clinically valid; parameter-randomization and label-randomization sanity checks, together with domain-expert review, are therefore required [20], [24], [35]. The framework is a prototype for research decision support and is not an approved medical device, and the predictions should not be reported as an autonomous diagnosis.

The right success criterion is therefore a calibrated tradeoff between privacy, utility, security and efficiency. A model that is marginally better but has a very impractical ciphertext footprint, poor privacy accounting, or inconsistent cross-hospital behaviour is not necessarily better. In contrast, extreme privacy settings that strip out clinically significant signal are not helpful. The evaluation in Table III is intended to highlight these trade-offs rather than obscure them behind a single accuracy score.

Conclusion

In this work, we proposed a defense-in-depth federated framework for explainable decision support for clinical applications. We use FedProx for non-IID optimisation, record-level DP-SGD for training-record influence constraint, CKKS for weighted aggregation in the ciphertext domain, and XAI for interpretation-facing to clinicians in the design. A key separated trust architecture limits decryption to the aggregate, as opposed to intentionally exposing individual client updates to the coordinator. It also describes the data splitting, ablation controls, privacy accounting, numerical error assessment in CKKS, communication profiling, threat boundary definitions and repeatability criteria. The desired contribution is a rigorous system integration and evaluation framework. Future works include deployment of the system on validated public clinical datasets, evaluation of threshold CKKS and defences against malicious clients, and quantification of the stability of XAI under privacy-preserving training.

References

1. E.-u.-H. Qazi, W. K. Al-Ghanem, M. H. Faheem, and H. Ullah, “Federated learning framework for privacy-preserving explainable AI-driven clinical decision-making,” IEEE J. Biomed. Health Inform., accepted version, 2026, doi: 10.1109/JBHI.2026.3679499.
2. N. Rieke et al., “The future of digital health with federated learning,” npj Digit. Med., vol. 3, Art. no. 119, 2020.
3. P. Kairouz et al., “Advances and open problems in federated learning,” Found. Trends Mach. Learn., vol. 14, nos. 1–2, pp. 1–210, 2021.
4. H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, “Communication-efficient learning of deep networks from decentralized data,” in Proc. AISTATS, 2017, pp. 1273–1282.
5. T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, “Federated optimization in heterogeneous networks,” in Proc. Mach. Learn. Syst., vol. 2, pp. 429–450, 2020.
6. N. T. Madathil, F. K. Dankar, M. Gergely, A. N. Belkacem, and S. Alrabaee, “Revolutionizing healthcare data analytics with federated learning: A comprehensive survey of applications, systems, and future directions,” Comput. Struct. Biotechnol. J., 2025.
7. R. Haripriya, N. Khare, and M. Pandey, “Privacy-preserving federated learning for collaborative medical data mining in multi-institutional settings,” Sci. Rep., vol. 15, no. 1, Art. no. 12482, 2025.
8. S. Gulati, K. Guleria, N. Goyal, A. A. Alzubi, and A. K. Castilla, “A privacy-preserving collaborative federated learning framework for detecting retinal diseases,” IEEE Access, 2024.
9. J. Mao, X. Ma, Y. Bi, and R. Zhang, “A comprehensive federated learning framework for diabetic retinopathy grading and lesion segmentation,” IEEE Trans. Big Data, 2024.
10. M. Al-Asfoor, M. H. Alsalihi, and K. Maher, “Brain tumor classification based on federated learning,” in Proc. 10th Int. Conf. Optimization and Applications (ICOA), 2024, pp. 1–4.
11. Z. U. Rehman, S. N. M. Shah, U. Hassan, and S. U. Islam, “Enhancing brain tumor classification using CNN: Impact of data augmentation on model generalization and performance in federated learning,” in Proc. 26th Int. Multi-Topic Conf. (INMIC), 2024, pp. 1–6.
12. M. Asad et al., “Limitations and future aspects of communication costs in federated learning: A survey,” Sensors, vol. 23, no. 17, Art. no. 7358, 2023.
13. M. F. Almufareh, N. Tariq, M. Humayun, and B. Almas, “A federated learning approach to breast cancer prediction in a collaborative learning framework,” Healthcare, vol. 11, Art. no. 3185, 2023.
14. J. H. Cheon, A. Kim, M. Kim, and Y. Song, “Homomorphic encryption for arithmetic of approximate numbers,” in Advances in Cryptology–ASIACRYPT, 2017, pp. 409–437.
15. C. Zhang, S. Li, J. Xia, W. Wang, F. Yan, and Y. Liu, “BatchCrypt: Efficient homomorphic encryption for cross-silo federated learning,” in Proc. USENIX ATC, 2020, pp. 493–506.
16. K. Bonawitz et al., “Practical secure aggregation for privacy-preserving machine learning,” in Proc. ACM CCS, 2017, pp. 1175–1191.
17. C. Dwork and A. Roth, The Algorithmic Foundations of Differential Privacy. Hanover, MA, USA: Now Publishers, 2014, vol. 9, nos. 3–4, pp. 211–407.
18. M. Abadi et al., “Deep learning with differential privacy,” in Proc. ACM CCS, 2016, pp. 308–318.
19. R. Shokri, M. Stronati, C. Song, and V. Shmatikov, “Membership inference attacks against machine learning models,” in Proc. IEEE Symp. Security and Privacy, 2017, pp. 3–18.
20. R. R. Selvaraju et al., “Grad-CAM: Visual explanations from deep networks via gradient-based localization,” in Proc. IEEE ICCV, 2017, pp. 618–626.
21. M. T. Ribeiro, S. Singh, and C. Guestrin, “Why should I trust you? Explaining the predictions of any classifier,” in Proc. ACM SIGKDD, 2016, pp. 1135–1144.
22. S. M. Lundberg and S.-I. Lee, “A unified approach to interpreting model predictions,” in Adv. Neural Inf. Process. Syst., vol. 30, 2017.
23. I. Y. Abushawish, S. Modak, E. Abdel-Raheem, S. A. Mahmoud, and A. J. Hussain, “Deep learning in automatic diabetic retinopathy detection and grading systems: A comprehensive survey and comparison of methods,” IEEE Access, vol. 12, pp. 84785–84802, 2024.
24. K. A. Alavee et al., “Enhancing early detection of diabetic retinopathy through the integration of deep learning models and explainable artificial intelligence,” IEEE Access, vol. 12, pp. 73950–73969, 2024.
25. E. Kotei and R. Thirunavukarasu, “A comprehensive review on advancement in deep learning techniques for automatic detection of tuberculosis from chest X-ray images,” Arch. Comput. Methods Eng., vol. 31, no. 1, pp. 455–474, 2024.

26. C. Liu, Y. Luo, Y. Xu, and B. Du, “Foundation models matter: Federated learning for multi-center tuberculosis diagnosis via adaptive regularization and model-contrastive learning,” World Wide Web, vol. 27, no. 3, Art. no. 30, 2024.
27. C. Liu, Y. Luo, Y. Xu, and B. Du, “HPFL: Hyper-network guided personalized federated learning for multi-center tuberculosis chest X-ray diagnosis,” Multimedia Tools Appl., vol. 83, no. 29, pp. 73273–73287, 2024.
28. P. K. Das, S. Sreevatsav, and A. Abraham, “An efficient deep learning network with orthogonal SoftMax layer for automatic detection of tuberculosis,” Eng. Appl. Artif. Intell., vol. 133, Art. no. 108116, 2024.
29. M. Faiyazuddin et al., “The impact of artificial intelligence on healthcare: A comprehensive review of advancements in diagnostics, treatment, and operational efficiency,” Health Sci. Rep., vol. 8, no. 1, Art. no. e70312, 2025.
30. M. Butt et al., “A fog-based privacy-preserving federated learning system for smart healthcare applications,” Electronics, vol. 12, no. 19, Art. no. 4074, 2023.
31. A. S. Budhewar, P. G. Patil, and S. M. Kale, “Neighbour-Aware Cooperation For Semi-Supervised Decentralized Machine Learning,” Educational Administration: Theory and Practice, vol. 30, no. 5, pp. 2039–2047, May 2024.
32. S. Budhewar, P. G. Patil, M. J. Patel, A. V. Roy, N. E. Phapale, and D. S. Rathod, “Secure Care Hub: A Blockchain-Enabled Platform for Streamlined Healthcare Services,” International Journal of Advanced Scientific Research and Engineering Trends, vol. 8, no. 9, pp. 11–14, 2024, doi: 10.65521/ijasret.v8i9.2332.
33. S. Budhewar et al., “Federated AI-Driven Digital Twins in the Healthcare Metaverse: Architectures, Privacy, and Clinical Intelligence,” in The Convergence of the Metaverse, AI, and Federated Learning in Healthcare, IGI Global Scientific Publishing, 2026, pp. 217–250, doi: 10.4018/979-8-3373-8889-2.ch008.
34. Mironov, “Renyi differential privacy,” in Proc. IEEE 30th Comput. Security Foundations Symp. (CSF), 2017, pp. 263–275.
35. J. Adebayo et al., “Sanity checks for saliency maps,” in Adv. Neural Inf. Process. Syst., vol. 31, 2018.