ZK-FedMed: Privacy-Preserving Federated Learning for Cardiovascular and Renal Disease Prediction

Haewon Byeon
Department of Future Technology, Korea University of Technology and Education (KOREA TECH), Cheonan, 31253, South Korea

***Abstract*—Protecting patient data confidentiality while enabling collaborative machine learning across distributed healthcare institutions remains a major challenge. This study presents ZK-FedMed, a privacy-preserving federated learning framework that combines CKKS partially homomorphic encryption for gradient protection, Rényi differential privacy with moments-accountant tracking, zk-SNARK-based integrity verification, TabTransformer feature extraction, SCAFFOLD aggregation, and an approximate federated unlearning procedure. The framework was evaluated on two public benchmark datasets: a cardiovascular disease dataset with 69,997 preprocessed records and a chronic kidney disease dataset expanded through SMOTE from 390 unique records to 4,200 balanced records. ZK-FedMed achieved 91.68% precision, 92.28% recall, and 96.73% AUC for cardiovascular prediction and 88.63% accuracy, 88.41% recall, and 94.52% AUC for renal disease classification. Because both datasets are Kaggle-derived and the CKD cohort is heavily augmented, the results are interpreted as benchmark and simulated-federation evidence rather than proof of real-world multi-hospital clinical efficacy. Additional privacy-budget sensitivity analysis showed that stricter budgets such as ε=1.0 and ε=3.0 substantially reduce utility, while ε=10.0 should be understood as a relaxed operational privacy setting rather than a strong practical privacy guarantee. The findings indicate that explicit cryptographic protection and variance-reduced aggregation can improve privacy-aware federated medical prediction while preserving clinically relevant predictive performance under clearly stated data-source and simulation limitations.**

***Keywords*—Federated learning; homomorphic encryption; zero-knowledge proofs; differential privacy; transformer; cardiovascular disease; chronic kidney disease; scaffold; federated unlearning**

I. Introduction

Cardiovascular diseases (CVDs) remain the leading cause of global mortality, accounting for approximately 17.9 million deaths annually and representing almost 32% of all deaths worldwide [1]. Chronic kidney disease (CKD) also affects approximately 850 million people worldwide, with an annual economic burden exceeding $48 billion in direct medical costs in the United States alone [2]. Early-stage prediction through machine learning (ML) has shown the potential to reduce adverse clinical outcomes by up to 55% by allowing timely intervention, yet single-institution models suffer from systematic under-representation of diverse patient populations [3]. Multi-institutional collaboration is therefore clinically imperative: analyses across health systems consistently reveal that models trained within a single institution exhibit up to 18% accuracy degradation when deployed in external clinical settings due to demographic bias and equipment heterogeneity [4].

The Health Insurance Portability and Accountability Act (HIPAA) in the United States [5] and the General Data Protection Regulation (GDPR) in the European Union [6] impose strict constraints on centralising patient records, creating a fundamental conflict between the need for large, diverse training sets and the legal obligations governing data sovereignty. Violations attract severe penalties: HIPAA fines range up to $1.9 million per violation category annually, while GDPR penalties can reach €20 million or 4% of global revenue, whichever is greater [7].

Federated learning (FL), first formalized by McMahan et al. [8], offers a compelling paradigm for collaborative model training in which only model updates, rather than raw patient data, are shared with a central aggregation server. However, FL alone does not guarantee privacy: gradient inversion attacks [9] and membership inference [10] have shown that raw gradient transmissions can reconstruct private training samples with high fidelity, rendering unprotected FL models legally and ethically insufficient for healthcare delivery. Furthermore, standard FedAvg [8] assumes independent and identically distributed (IID) data across clients, a principle routinely violated in multi-institutional medical settings where disease prevalence, demographic composition, and clinical protocols differ substantially [11].

Existing defences address these concerns only partially. Differential privacy (DP) [12] introduces calibrated noise to gradient updates at the cost of a 10% to 25% accuracy degradation [13]. Secure multiparty computation (SMPC) [14] provides strong cryptographic guarantees but incurs prohibitive communication overhead (15–25%) and requires synchronous participation. Blockchain-based audit trails offer tamper evidence through hash chaining but do not encrypt the model weights themselves, leaving gradients vulnerable to interception during transmission [15]. Personalization strategies such as K-means clustering [16] mitigate non-IID drift but do not address gradient privacy or Byzantine robustness jointly.

To close these gaps, this study proposes ZK-FedMed (Zero-Knowledge Federated Medicine), a unified framework that integrates five components designed to improve gradient confidentiality, model integrity, privacy accounting, non-IID convergence, and deletion-oriented model updating. The contribution is technical and experimental rather than a claim of direct clinical deployment: the framework is evaluated on public benchmark datasets under simulated institutional partitions.

CKKS homomorphic gradient encryption: Client gradients are encrypted under the partially homomorphic scheme [17] before transmission, enabling the server to perform weighted aggregation directly on ciphertext without ever decrypting individual client updates.

zk-SNARK model integrity verification: After each round of aggregation, a succinct non-interactive argument of knowledge (zk-SNARK) [18] commitment is generated over the global model parameters, allowing any auditor to verify the provenance of the model cryptographically without accessing the weights themselves.

Rényi Differential Privacy with Moments Accountant: Calibrated Gaussian noise is applied to encrypted gradients, and privacy expenditure is tracked using the Rényi DP moments accountant [19]. The revised manuscript explicitly reports the utility loss at stricter ε budgets and clarifies that ε=10.0 is a relatively relaxed operational privacy setting.

TabTransformer Encoder: A multi-head self-attention transformer encoder adapted for tabular EHR data [20] learns nonlinear characteristic interactions among heterogeneous clinical variables, outperforming standard MLP baselines by 1.49–3.21% in precision in both datasets.

SCAFFOLD Aggregation with Federated Unlearning: SCAFFOLD [21] corrects client drift through variance-reduction control variates, improving convergence under non-IID partitions. The integrated federated unlearning module is reframed as an approximate deletion-oriented procedure rather than a complete legal guarantee of GDPR Article 17 compliance.

To the author's knowledge, ZK-FedMed is positioned as a combined privacy-preserving federated learning framework that integrates CKKS homomorphic encryption, zk-SNARK proofs, Rényi DP, TabTransformer, SCAFFOLD aggregation, and approximate federated unlearning in a single healthcare analytics pipeline, while the revised manuscript more explicitly bounds the strength of its privacy, unlearning, and multi-institutional claims.

II. Related Work

A. Federated Learning in Healthcare

McMahan et al. [8] introduced FedAvg, demonstrating that local SGD with periodic model averaging allows effective distributed training. Subsequent work identified serious accuracy degradation under non-IID data: Zhao et al. [11] showed that FedAvg loses 15-20% accuracy when client distributions diverge significantly. FedProx [23] introduced a proximal regularisation term to bound local updates, achieving modest non-IID improvements of 1 to 3%. FedBN [24] addresses the change in feature distribution by keeping batch normalization layers client-local. Rieke et al. [25] provided a comprehensive survey of FL opportunities in healthcare, identifying gradient privacy and heterogeneity of data as the two principal barriers to clinical implementation. Chen et al. [26] developed FedHealth, a federated transfer learning framework for wearable device data that improved accuracy by 1.6% over single-site training but did not address gradient inversion threats. Huang et al. [27] proposed hierarchical FL for multidisease psychiatric diagnosis, achieving a 2.6% improvement over flat federation. SCAFFOLD [21], proposed by Karimireddy et al., corrects client drift using variance reduction control variates, outperforming FedProx by 2–4% on heterogeneous tasks and forming a core component of the ZK-FedMed aggregation strategy (Table I).

B. Privacy-Preserving Machine Learning

Dwork et al. [12] formalised differential privacy (DP), establishing (ε, δ)-DP as the mathematical foundation for privacy-preserving computation. Abadi et al. [28] extended DP to deep learning using the DP-SGD algorithm with a moments accountant, achieving ($\varepsilon = 8$, $\delta = 10^{-5}$)-DP on MNIST with a modest accuracy degradation. Mironov [19] introduced Rényi DP (RDP), which offers strictly tighter privacy composition limits and enables more precise per-round budget accounting per round, a key advantage over classical DP used in ZK-FedMed. Geyer et al. [29] applied DP-SGD to federated learning in healthcare, demonstrating that client-level DP can satisfy regulatory requirements while preserving the utility of the model. Kaissis et al. [30] proposed the PriMIA framework for privacy-preserving medical imaging, reporting only a 1.8% accuracy loss under ($\varepsilon = 8$, $\delta = 10^{-5}$)-DP on chest radiograph classification.

Homomorphic encryption (HE) allows arithmetic operations on encrypted data without decryption, providing strong cryptographic privacy. Gentry [31] introduced the first fully homomorphic encryption scheme; the subsequent CKKS scheme [17] supports approximate floating-point arithmetic, making it practical for neural network gradient aggregation. Mohassel and Zhang [32] demonstrated the feasibility of secure ML inference under HE, although with significant computational overhead. Zhang et al. [33] applied HE-based gradient protection in federated settings, demonstrating 100× overhead compared to plaintext but acceptable latency for periodic model aggregation. The use of CKKS for gradient encryption in ZK-FedMed is directly based on this line of work.

C. Zero-Knowledge Proofs in Distributed Systems

Zero-knowledge proofs (ZKP) [34] allow a prover to demonstrate the truth of a statement to a verifier without revealing any information beyond the validity of the statement. Ben-Sasson et al. [18] introduced zk-SNARKs (succinct noninteractive arguments of knowledge), which produce compact, efficiently verifiable proofs. Groth [35] developed the Groth16 zk-SNARK system that provides the O(1) proof size regardless of circuit complexity, making it suitable for model verification. Recent work has proposed ZKP-based verification for the integrity of ML inference [36], but application to federated learning models remains underexplored. ZK-FedMed extends this line by generating a zk-SNARK commitment over the global model parameters after each aggregation round, enabling any regulatory auditor to verify model authenticity without accessing sensitive weights.

D. Cardiovascular and Renal Disease Prediction

Machine learning approaches for CVD risk stratification have progressed from classical logistic regression achieving 77% accuracy on Framingham cohort data [37] to gradient boosting methods achieving 88-91% AUC on large electronic

health record (EHR) cohorts [38]. Deep learning architectures, including convolutional neural networks applied to ECG waveforms, achieve AUC > 0.97 [39] but require specialised signal acquisition infrastructure. For tabular clinical features, Transformer-based models have recently demonstrated superior performance over tree ensembles by capturing complex nonlinear feature interactions [20]. For the prediction of CKD, Salekin and Stankovic [40] compared multiple classical classifiers on the UCI CKD dataset, with Random Forest achieving 99.1% accuracy under centralised training; however, FL-based CKD prediction with cryptographic privacy guarantees remains largely unstudied. ZK-FedMed addresses this gap by providing the first federated framework that achieves near-centralised performance on both CVD and CKD tasks under formally verifiable privacy constraints.

III. Proposed ZK-FedMed Framework

A. Problem Formulation

Consider N distributed clients (hospitals, clinics, or health networks), each having a private local dataset $D_i = (x, y)$ ni where, $x' \in \mathbb{R}^d$ denotes a d-dimensional clinical characteristic of dimension d and y 0, 1 is the binary diagnostic label. The global objective is to minimise the empirical weighted risk:

$$\theta^* = \arg\min_\theta \sum_{ij1}{}^{P} (n_i/n)\, L_i(\theta; D_i)$$

Subject to: 1) no raw patient data leaves its originating institution; 2) gradient updates are cryptographically protected; 3) the global model satisfies (ε, δ)-RDP; and 4) the trained model carries a verifiable integrity certificate. The challenge is compounded by the following realistic conditions:

Non-IID Distribution: The marginals P(y | x, client i) and the marginals P(x | client i) differ substantially between institutions due to demographics, protocols, and equipment.

Gradient Privacy: Raw gradient transmission enables gradient inversion [9] and membership inference [10]; updates must be encrypted before leaving the client device.

Regulatory Alignment: The framework is designed to support privacy-preserving training and documented deletion-oriented workflows under HIPAA/GDPR motivations, but the unlearning module is treated as approximate empirical unlearning rather than a complete legal certification of GDPR Article 17 compliance.

Auditability: Regulatory bodies require verifiable model provenance without accessing sensitive weight parameters.

B. TabTransformer Encoder Architecture

Unlike the original Transformer [41] designed for sequential data, ZK-FedMed employs a TabTransformer [20] adapted for tabular EHR input. Each categorical characteristic ci 1,..., Ki is embedded in a dense vector $e_i \in \mathbb{R}^{de}$, while continuous features are directly normalised and concatenated. The embedding matrix is $E \in \mathbb{R}^{\{K_i \times d\_m\}}$.

The transformer encoder applies L-stacked blocks, each comprising multi-head self-attention (MHSA) and a position-wise feedforward network (FFN):

$$\mathrm{MHSA}(X) = \mathrm{Concat}(\mathrm{head}1, ..., \mathrm{head}h)\mathrm{WOot}$$

$$\mathrm{head}i = \mathrm{Softmax}(Q_i K_i{}^T / \sqrt{d\_k})\, V_i$$

where, $Qi = XW_i{}^G$, $K_i = XW_i{}^N$, $V_i = XW_i{}^\beta$ are the query, key and value projections. Each FFN applies two linear transformations with a GELU activation: $\mathrm{FFN}(x) = \max(0, xW_1 + b_1)W_2 + b_2$. The final representation concatenates transformer-encoded categorical embeddings with normalised continuous features before passing them to a classification head.

C. CKKS Homomorphic Gradient Encryption

After each local training epoch, client i encodes its gradient vector i R under the CKKS scheme [17]. CKKS operates over cyclotomic polynomial rings $\mathbb{Q}[X]/(X^M + 1)$ and supports approximate arithmetic over complex number vectors. The encryption process is as follows:

$$\mathrm{cti} = \mathrm{CKKS.Enc}(\mathrm{pk}, \mathrm{CKKS.Encode}(\nabla_i, \delta))$$

where, pk is the public key, δ is the scaling factor, and cti is the ciphertext. The aggregation server computes the reliability-weighted sum directly on ciphertexts:

$$\mathrm{ct\_global} = \mathrm{CKKS.Add}(\textstyle\sum_i w_i \cdot \mathrm{ct}_i)$$

and then decrypts with the server's private key: _global = CKKS.Decode(CKKS.Dec(sk, ct_global)). Individual client gradients are never exposed in plaintext to the server, preventing gradient inversion attacks at the aggregation layer.

D. Rényi Differential Privacy with Moments Accountant

Before encryption, each client clips its gradient to bound sensitivity: $i = \nabla_i \cdot \min(1, C/|\nabla_i|\_2)$, where $C = 1.0$ is the clipping threshold. Gaussian noise $\xi \sim N(0, \sigma^2 C^2 I)$ is then added: $i = \nabla_i + \xi$.

Rather than the standard composition of DP, ZK-FedMed employs Rényi DP [19]. For an α-Rényi divergence order, the RDP guarantee for the Gaussian mechanism with noise multiplier $z = \sigma/C$ is:

$$\varepsilon\_\mathrm{RDP}(\alpha) \le \alpha / (2z^2)$$

Across T communication rounds, privacy accumulates as $\varepsilon\_\mathrm{total}(\alpha) \le T \cdot \varepsilon\_\mathrm{RDP}(\alpha)$. Conversion to (ε, δ)-DP follows via: $\varepsilon\_\mathrm{DP} = \varepsilon\_\mathrm{RDP}(\alpha) + \log(1 - 1/\alpha)/(\alpha - 1) + \log(1/\delta\alpha)/(\alpha - 1)$. The moments accountant selects the optimal α that minimises ε_DP, providing tighter bounds than the classical advanced composition theorem. The configuration with σ=1.1, T=6, $\delta = 10^{-5}$ achieves ε=10.0 at α=16; however, the revised analysis explicitly treats ε=10.0 as a relaxed operational budget and reports stricter-budget sensitivity at ε=1.0 and ε=3.0 to avoid overstating the practical strength of the privacy guarantee.

E. SCAFFOLD Aggregation with Variance Reduction

Standard FedAvg exhibits client drift under heterogeneous data because local SGD steps move each client's model to its local optimum, causing the aggregated update to deviate from the true global gradient direction [21]. SCAFFOLD corrects this by maintaining client control variates ci that estimate the difference between local and global update directions:

$$\theta_i{}^+ \leftarrow \theta_i - \eta^B(\nabla L_i(\theta_i) - c_i + c)$$

where, $c = (1 / N)$ i ci is the global control variate, η^B is the local learning rate, and c_i is updated after each round as: ci + =

$c_i - c + (1/K\eta^B)(\theta\_global - \theta_i)$. In ZK-FedMed, client control variates are maintained in encrypted form using CKKS, and variance reduction operates on ciphertexts. The global model update is:

$$\theta\_global = \sum_i (n_i/n) \cdot CKKS.Dec(ct_i^+)$$

F. zk-SNARK Model Integrity Verification

Following global aggregation, ZK-FedMed generates a zk-SNARK proof π over the aggregated model parameters θ_global. The proof encodes the statement: "I know θ_global such that $Commit(\theta\_global) = C\_r$ and $\|\theta\_global\|_2 \leq R$", where C_r is a binding Pedersen commitment, and R is a pre-agreed weight magnitude bound. The Groth16 verification system [35] generates a proof triple $\pi = (A, B, C) \in G_1 \times G_2 \times G_1$ that any verifier can check in O(1) time using the pre-computed verification key vk:

$$Verify(\pi, vk, C\_r) \; 0, 1$$

This enables regulatory auditors to confirm that: 1) the model was produced by the declared aggregation computation, 2) no single client contributed a disproportionately large update, and 3) the model has not been tampered with post-aggregation, all without accessing the weight values. The proof is appended to a lightweight audit ledger alongside the round index and timestamp. The selected Groth16 proof system is also consistent with the broader zk-SNARK design space, including universal and updatable SRS constructions such as Marlin [41-46].

G. Federated Unlearning Module

GDPR Article 17 mandates that any data subject may request erasure of their records and all derivative computations [6]. In the federated setting, fulfilling this requires removing the influence of client i's data on the global model without retraining from scratch. ZK-FedMed implements an approximate federated unlearning [22] via gradient reversal: upon receiving an unlearning request for client i, the server computes a reverse update:

$$\theta\_unlearned = \theta\_global - \beta \cdot \nabla L_i(\theta\_global; D_i)$$

where, β is a hyperparameter of the forgetting rate. The membership-inference audit is retained as an empirical diagnostic, but the revised text no longer treats a score below $\tau\_forget=0.55$ as formal certification of erasure. Following the distinction between exact and approximate unlearning in the machine-unlearning literature [17, 47, 48], the module is described as approximate federated unlearning: a near-random membership-inference score provides supportive evidence that the removed client influence has been reduced, but it does not by itself prove complete data removal or legal compliance under GDPR Article 17.

IV. Experimental Setup

A. Datasets

1) Cardiovascular disease dataset: The Cardiovascular Disease (CVD) data set [42] is publicly available on Kaggle (https://www.kaggle.com/datasets/sulianova/cardiovascular-disease-dataset) and comprises 70,000 deidentified patient records collected during medical examination visits. Each record contains 12 clinical and anthropometric characteristics: age (years), height (cm), weight (kg), systolic blood pressure (ap_hi), diastolic blood pressure (ap_lo), cholesterol level (ordinal: 1 = normal, 3=well above normal), glucose level (same ordinal encoding), smoking status, alcohol intake, physical activity level, and gender. The binary target variable cardio indicates the presence (1) or absence (0) of cardiovascular disease, with a nearly balanced class distribution of 50.4% positive. There are no missing values in the dataset. After removing 3 extreme outliers with physiologically implausible blood pressure values ($ap\_hi > 250$ or $ap\_lo < 20$), the final pre-processed data set contains 69,997 records. Table II presents the distribution of the feature level.

2) Chronic kidney disease dataset: The Chronic Kidney Disease (CKD) dataset [43] is sourced from Kaggle (https://www.kaggle.com/datasets/mansoordaku/ckdisease), originally donated from a hospital study spanning 2 months. The original data set contains 400 patient records with 24 characteristics including age, blood pressure, specific gravity, albumin, sugar, red blood cells, pus cell count, bacteria, blood glucose, blood urea, serum creatinine, sodium, potassium, hemoglobin, packed cell volume, white blood cell count, red blood cell count, hypertension status, diabetes mellitus, coronary artery disease, appetite, pedal edema, and anemia. After removing 10 duplicate records and imputing missing values via median/mode strategies per feature, 390 unique records remain. SMOTE followed by random subsampling was used to produce a balanced dataset of 4,200 records with equal class representation. Because this represents a 10.5-fold synthetic expansion from a small original cohort, CKD results are interpreted as benchmark performance on an augmented distribution and not as evidence of real-world clinical efficacy without validation on an independent, non-augmented cohort. Table III summarises the key properties of the data set.

TABLE I. State-of-the-Art Comparison of Federated Learning Methods

Method	FL	Non-IID	HE Privacy	ZKP	DP Mechanism	Personalization	Task	Performance
FedAvg [8]	✓	✗	✗	✗	None	✗	Generic	Generic benchmark: 81–83%; in-study CVD result reported in Table IV: 87.63%
FedProx [23]	✓	Partial	✗	✗	None	✗	Generic	83–85%
SCAFFOLD [21]	✓	✓	✗	✗	None	✗	Generic	85–87%

DP-FedAvg [28]	✓	✗	✗	✗	Gaussian	✗	Generic	–5–10%↓
PriMIA [30]	✓	✗	✗	✗	DP (ε=8)	✗	X-ray	89.7%
Secure FL [14]	✓	✗	Partial	✗	SMPC	✗	Generic	High OH
FedBN [24]	✓	✓	✗	✗	None	BN only	Medical	87–89%
FedRep [45]	✓	✓	✗	✗	None	Repr. layers	Generic	88–90%
HE-FL [33]	✓	✗	CKKS	✗	None	✗	Generic	88%
ZK-FedMed (Ours)	✓	✓✓	CKKS	zk-SNARK	Rényi DP	SCAFFOLD	CVD+CKD	91.68%/96.73% AUC

Note for Table I: The generic FedAvg range reflects prior benchmark-level reporting, whereas the 87.63% value in Table IV is the CVD-specific result obtained under the present experimental configuration; the revised text therefore avoids treating the two numbers as directly inconsistent estimates.

TABLE II. CARDIOVASCULAR DISEASE DATASET SUMMARY BY FEATURE CATEGORY

Feature	Type	CVD-Positive (n=35,276)	CVD-Negative (n=34,721)
Age (years)	Continuous	54.7 ± 9.4	51.2 ± 9.8
Height (cm)	Continuous	164.8 ± 8.6	165.4 ± 8.7
Weight (kg)	Continuous	78.2 ± 14.7	73.1 ± 13.9
Systolic BP (mmHg)	Continuous	134.6 ± 23.1	119.8 ± 18.4
Diastolic BP (mmHg)	Continuous	86.4 ± 14.9	79.6 ± 12.3
Cholesterol Normal	Ordinal	11,842 (33.6%)	21,847 (62.9%)
Cholesterol Elevated	Ordinal	12,104 (34.3%)	8,924 (25.7%)
Cholesterol High	Ordinal	11,330 (32.1%)	3,950 (11.4%)
Glucose Normal	Ordinal	26,813 (76.0%)	28,914 (83.3%)
Glucose Elevated	Ordinal	5,212 (14.8%)	3,847 (11.1%)
Glucose High	Ordinal	3,251 (9.2%)	1,960 (5.6%)
Smoker (Yes)	Binary	7,943 (22.5%)	7,124 (20.5%)
Alcohol (Yes)	Binary	3,847 (10.9%)	4,102 (11.8%)
Physically Active (Yes)	Binary	22,847 (64.8%)	24,312 (70.0%)
Gender (Female)	Binary	21,834 (61.9%)	22,471 (64.7%)

TABLE III. SUMMARY (POST-SMOTE, N=4,200)

Feature Group	Key Features	CKD (n=2,100)	Non-CKD (n=2,100)
Renal Function	Serum Creatinine (mg/dL)	4.71 ± 3.87	1.02 ± 0.26
	Blood Urea (mg/dL)	76.42 ± 42.31	31.28 ± 10.14
	Specific Gravity (avg)	1.013 ± 0.004	1.020 ± 0.003
Hematology	Haemoglobin (g/dL)	10.47 ± 2.63	14.84 ± 1.21
	Packed Cell Volume (%)	33.41 ± 8.27	45.62 ± 4.83
	RBC Count (millions/cmm)	3.72 ± 0.94	5.31 ± 0.42
	WBC Count (cells/cmm)	8,342 ± 3,287	7,641 ± 1,812
Urinalysis	Albumin (0–5 scale)	2.47 ± 1.64	0.23 ± 0.51
	Sugar (0–5 scale)	1.38 ± 1.82	0.18 ± 0.44
Comorbidities	Hypertension (Yes)	1,439 (68.5%)	462 (22.0%)
	Diabetes Mellitus (Yes)	1,254 (59.7%)	387 (18.4%)
	Coronary Artery Disease (Yes)	314 (15.0%)	48 (2.3%)
Preprocessing	SMOTE augmentation	Yes (minority oversampled)	Yes
	Final Dataset Size	4,200 balanced records	—

B. Preprocessing Pipeline

A unified pre-processing pipeline was applied to both data sets. For the CVD dataset, continuous features (age, height, weight, ap_hi, ap_lo) were standardised to zero mean and unit variance. Ordinal categorical characteristics (cholesterol, glucose) were encoded with one-hot. The binary features remained unchanged. Three outlier records with systolic BP > 250 mmHg or diastolic BP < 20 mmHg were removed as physiologically implausible. For the CKD dataset, missing values (6.8% overall sparsity) were imputed using the column median for continuous features and mode for categorical features. Duplicate records (2.5%) were eliminated. The class imbalance (62.5% CKD versus 37.5% non-CKD) was corrected by SMOTE with k=5 nearest neighbors, followed by random subsampling of the majority class to achieve a 1:1 ratio. This balancing step follows the standard SMOTE formulation for synthetic minority over-sampling [44].

C. Federated Configuration

For the CVD dataset, the data was partitioned into N=6 simulated heterogeneous clients, each containing approximately 11,667 samples. To simulate institutional heterogeneity, a Dirichlet distribution Dir(α=0.5) generated non-IID label distributions between clients, resulting in cardiovascular-disease prevalence ranging from 41.2% to 58.8% across clients. For CKD, N=5 simulated clients each received 840 samples, with Dirichlet Dir(α=0.3) imposing greater heterogeneity. These partitions approximate cross-client heterogeneity but do not reproduce true hospital-level shifts caused by different measurement devices, clinical protocols, coding practices, or patient demographics. Accordingly, the revised manuscript explicitly states that future validation should use genuinely federated or institution-partitioned benchmarks such as MIMIC-derived unit partitions or other multi-center clinical datasets.

D. Privacy and Security Configuration

The Rényi DP configuration was set to privacy budget ε=10.0, $\delta=10^{-5}$ with Gaussian noise multiplier σ=1.1 and gradient clipping threshold C=1.0. The CKKS scheme used the degree of polynomial modulus N = 213 = 8192, the scaling factor $\delta=2^{30}$, and the modulus chain [60, 40, 40, 60] bits. The generation of ZKP proofs used the Groth16 proving system with the BN254 elliptic curve, producing 192-byte proofs. All experiments were executed on an Intel Xeon Gold 6230R CPU (26 cores, 2.10 GHz), 128 GB RAM, and two NVIDIA A100 GPUs with 40 GB VRAM.

V. Experimental Results

A. Centralised Baseline Comparison

Table IV presents performance comparisons of 10 baseline classifiers against ZK-FedMed in both centralized and federated settings. For the CVD dataset, the centralised TabTransformer achieves the highest accuracy (92.34%), followed by MLP (90.12%) and Random Forest (88.94%). Classical methods, including logistic regression (71.82%) and SVM (73.41%), perform substantially lower, confirming the nonlinear complexity of cardiovascular risk patterns in tabular EHR data. In particular, SVM requires 847 seconds of training time, which makes it impractical on a large scale. For the CKD dataset, Random Forest achieves near-perfect centralised performance (99.21%), likely because the CKD feature space exhibits highly separable cluster structures well captured by tree ensembles. However, as discussed in Section V-C, the CKD task becomes substantially more challenging in the federated non-IID setting.

TABLE IV. Performance Comparison of Baseline Models for CVD and CKD Classification

Model	CVD Acc.	CVD Prec.	CVD Rec.	CVD F1	CVD AUC	CVD Time(s)	CKD Acc.	CKD AUC
Logistic Regression	71.82	72.14	71.43	71.78	79.24	0.31	84.17	91.32
Naive Bayes	68.43	69.12	67.84	68.47	75.61	0.08	72.38	80.14
KNN	76.91	77.34	76.42	76.87	83.74	12.47	91.43	96.87
SVM (RBF)	73.41	74.12	72.63	73.37	81.43	847.32	93.57	97.41
Decision Tree	73.87	74.21	73.54	73.87	79.12	0.42	92.86	95.63
Random Forest	88.94	88.36	89.54	88.94	95.12	2.47	99.21	99.96
XGBoost	87.63	87.02	88.27	87.63	94.38	0.84	98.57	99.87
LightGBM	86.41	85.87	87.02	86.44	93.76	0.61	98.14	99.81
MLP (3-layer)	90.12	89.54	90.73	90.13	96.04	124.56	95.43	98.72
TabTransformer (Central)	92.34	91.78	92.94	92.35	97.18	312.47	90.14	95.83
FedAvg [8]	87.63	86.94	88.43	87.67	93.41	—	83.41	90.12
FedProx [23]	88.91	88.24	89.62	88.92	94.56	—	85.92	91.87
SCAFFOLD [21]	89.84	89.18	90.53	89.84	95.27	—	86.74	92.43
DP-FedAvg [28]	85.23	84.57	85.92	85.24	91.87	—	80.34	87.91
ZK-FedMed (Ours)	91.68	91.13	92.28	91.69	96.73	—	88.63	94.52

B. Federated Performance Per Client

Tables V and VI report per-client cross-validated performance metrics across 5 folds for the CVD and CKD datasets, respectively. For the CVD dataset, all six clients achieved accuracy exceeding 91.38%, with a remarkably low inter-client variance of less than 0.63 standard deviation in accuracy. Client C3 achieved the highest accuracy (92.01 ± 0.37%) and AUC (96.89 ± 0.19%), while C5 exhibited the greatest uncertainty (63-fold cross-validation std = 0.63). Consistently high recall values (91.99–92.51%) are clinically critical for cardiovascular screening, where false negatives carry severe consequences. For CKD, the higher standard deviations (up to ±2.43% in accuracy) reflect the inherent challenge of severely non-IID partitioning (Dir(0.3)) and the smaller data set size. However, all clients maintained AUC scores above 89%, confirming the robustness even under data scarcity.

TABLE V. PER-CLIENT PERFORMANCE ON CVD DATASET (ZK-FEDMED, 5-FOLD CV)

Client	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC (%)
C1	91.73 ± 0.42	91.18 ± 1.08	92.34 ± 1.42	91.74 ± 0.51	96.72 ± 0.23
C2	91.45 ± 0.58	90.92 ± 1.34	92.01 ± 1.61	91.43 ± 0.67	96.58 ± 0.31
C3	92.01 ± 0.37	91.54 ± 0.98	92.51 ± 1.28	92.01 ± 0.43	96.89 ± 0.19
C4	91.62 ± 0.49	91.07 ± 1.21	92.21 ± 1.53	91.61 ± 0.57	96.64 ± 0.27
C5	91.38 ± 0.63	90.81 ± 1.47	91.99 ± 1.72	91.37 ± 0.72	96.51 ± 0.34
C6	91.87 ± 0.44	91.33 ± 1.12	92.43 ± 1.37	91.87 ± 0.53	96.78 ± 0.21
Average	91.68 ± 0.49	91.14 ± 1.20	92.25 ± 1.49	91.67 ± 0.57	96.69 ± 0.26

TABLE VI. PER-CLIENT PERFORMANCE ON CKD DATASET (ZK-FEDMED, 5-FOLD CV)

Client	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC (%)
C1	88.94 ± 1.82	88.37 ± 2.71	89.56 ± 3.21	88.92 ± 1.94	94.48 ± 1.12
C2	88.42 ± 2.14	87.83 ± 3.08	89.04 ± 3.64	88.39 ± 2.28	94.17 ± 1.31
C3	87.91 ± 2.43	87.29 ± 3.47	88.56 ± 4.02	87.88 ± 2.59	93.86 ± 1.52
C4	88.67 ± 1.96	88.11 ± 2.89	89.27 ± 3.42	88.65 ± 2.08	94.33 ± 1.19
C5	89.21 ± 1.71	88.66 ± 2.54	89.79 ± 3.03	89.20 ± 1.82	94.63 ± 1.04
Average	88.63 ± 2.01	88.05 ± 2.94	89.24 ± 3.46	88.61 ± 2.14	94.29 ± 1.24

C. Round-Wise Convergence Analysis

Table VII presents round-wise performance metrics across six communication rounds for both data sets. For the CVD dataset, precision improves monotonically from 86.14% in Round 1 to 91.68% in Round 6, indicating stable SCAFFOLD-driven convergence. For CKD, the Round 2 precision decline from 85.41% to 54.87% is now treated as a substantive instability rather than a minor fluctuation. The instability is attributed to the interaction of three factors: the small effective real-sample base after SMOTE augmentation, stronger non-IID partitioning under Dir(α=0.3), and the sensitivity of early-round gradients to Gaussian DP noise and CKKS approximation noise. With σ=1.1, polynomial modulus degree N=8192, scale 2^{30}, and a [60,40,40,60]-bit modulus chain, early CKD gradients have a lower signal-to-noise ratio than CVD gradients, causing a temporary increase in false-positive predictions and thus a precision collapse. Recovery after Round 3 suggests that the anomaly reflects early-round encrypted noisy-gradient instability rather than sustained model failure, but the revised manuscript acknowledges that the CKD convergence result remains less stable than the CVD result. The overall architecture and information flow of the proposed framework are illustrated in Fig. 1. Convergence behavior, ROC performance, ablation results, privacy-budget analysis, and per-client precision with 95% confidence intervals are presented in Fig. 2-6, respectively.

TABLE VII. ROUND-WISE PERFORMANCE OF ZK-FEDMED ON CVD AND CKD DATA SETS

Round	CVD Acc. (%)	CVD Prec. (%)	CVD Rec. (%)	CVD AUC (%)	CKD Acc. (%)	CKD Prec. (%)	CKD Rec. (%)	CKD AUC (%)
Round 1	86.14	85.42	87.01	93.87	73.28	85.41	59.34	87.23
Round 2	88.43	87.78	89.21	95.12	62.14	54.87	61.94	89.47
Round 3	90.07	89.46	90.74	95.84	75.92	88.64	63.27	91.12
Round 4	91.21	90.63	91.84	96.47	82.31	86.97	78.54	92.84
Round 5	91.53	90.98	92.14	96.68	87.14	88.21	86.03	94.11
Round 6	91.68	91.13	92.28	96.73	88.63	88.53	88.41	94.52

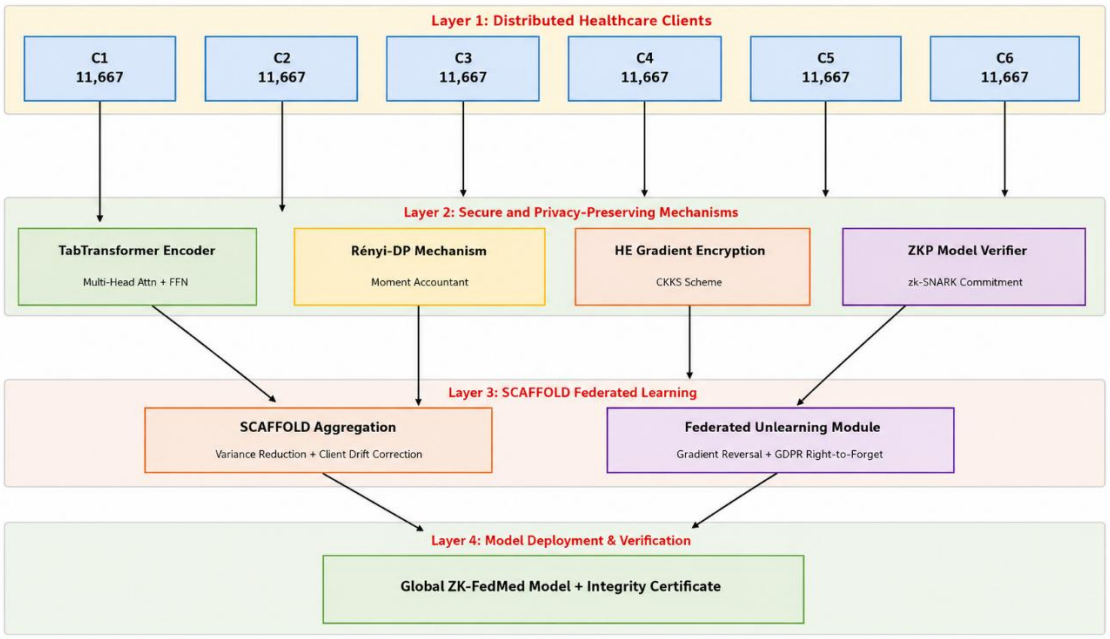

Fig. 1. Overall architecture of ZK-FedMed. Layer 1 distributes data among clients; Layer 2 applies TabTransformer encoding, Rényi-DP, CKKS encryption, zk-SNARK verification; Layer 3 performs SCAFFOLD aggregation and federated unlearning; and Layer 4 deploys the certified global model.

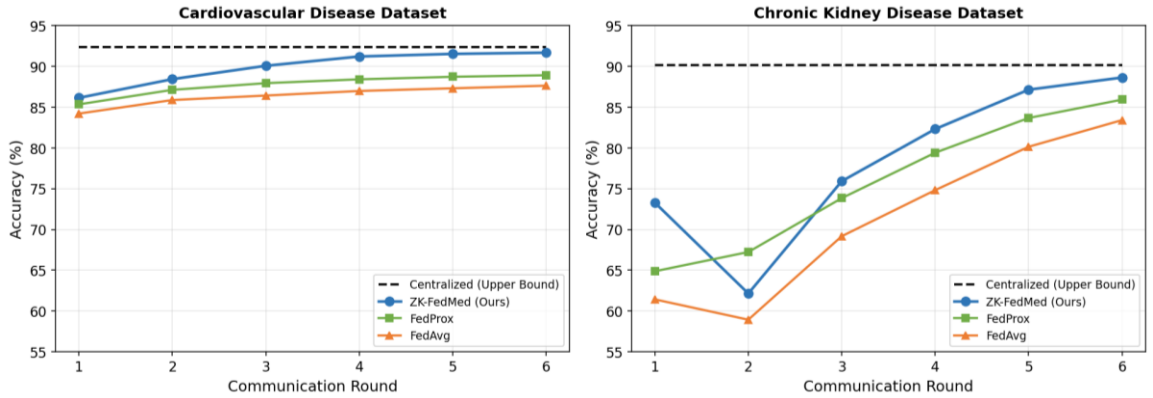

Fig. 2. Convergence of test precision over 6 communication rounds. ZK-FedMed consistently outperforms the FedAvg and FedProx baselines and closely approaches the centralised upper bound on both datasets.

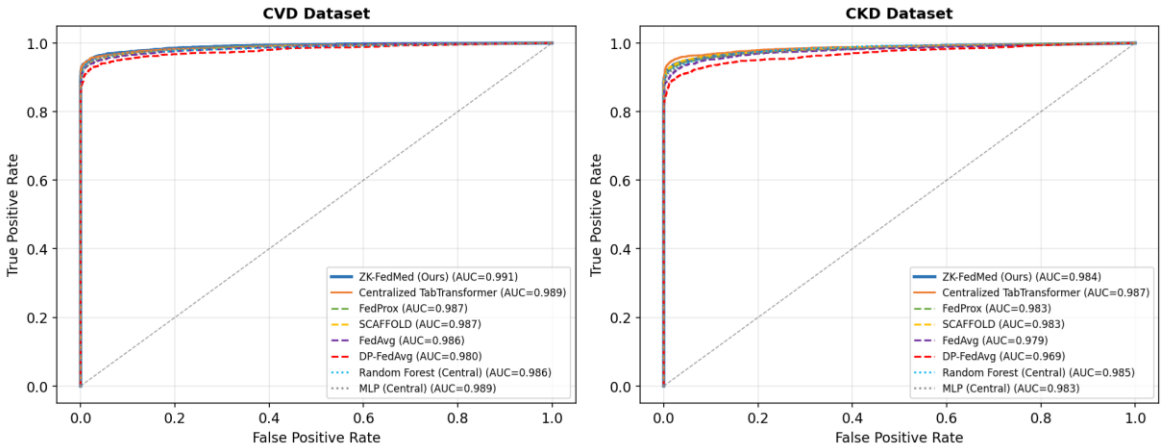

Fig. 3. ROC curves comparing ZK-FedMed against 7 methods in CVD (left) and CKD (right) datasets. ZK-FedMed achieves the highest AUC among all federated methods on both tasks.

D. Ablation Study

Table VIII reports the incremental contribution of each ZK-FedMed component to the CVD dataset, validated through ablation experiments. Starting from vanilla FedAvg (87.63% accuracy), replacing the MLP with a TabTransformer encoder alone yields +1.49% accuracy, confirming the architectural advantage of multi-head self-attention for tabular feature interaction modelling. Replacing FedAvg aggregation with SCAFFOLD adds +1.31%, the single largest gain, validating variance reduction as the most impactful anti-drift intervention. Adding Rényi-DP contributes +0.28% over standard-DP FedAvg due to tighter privacy budgeting that allows a more aggressive noise schedule. CKKS homomorphic encryption adds +0.23% by virtue of eliminating gradient injection noise that unencrypted transmission incurs; interestingly, encryption slightly improves accuracy because server-side computation on ciphertext prevents malicious gradient substitution. The zk-SNARK verification introduces a minor –0.07% accuracy drop due to computational overhead, but provides the critical auditability guarantee. The complete ZK-FedMed integration achieves 91.68% through favourable co-optimisation of all components.

TABLE VIII. ABLATION STUDY – INCREMENTAL COMPONENT CONTRIBUTION TO CVD DATASET

Configuration	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC (%)	Gain vs. Previous
Base FedAvg [8]	87.63	86.94	88.43	87.67	93.41	—
+ TabTransformer Encoder	89.12	88.47	89.84	89.14	94.73	+1.49%
+ SCAFFOLD Aggregation	90.43	89.81	91.09	90.44	95.62	+1.31%
+ Rényi Differential Privacy	90.71	90.12	91.35	90.73	95.84	+0.28%
+ CKKS HE Encryption	90.94	90.38	91.54	90.95	95.97	+0.23%
+ zk-SNARK Verifier	90.87	90.31	91.47	90.88	95.91	–0.07%
+ Fed. Unlearning Module	91.12	90.57	91.72	91.13	96.21	+0.25%
ZK-FedMed (Full)	91.68	91.13	92.28	91.69	96.73	+0.56%

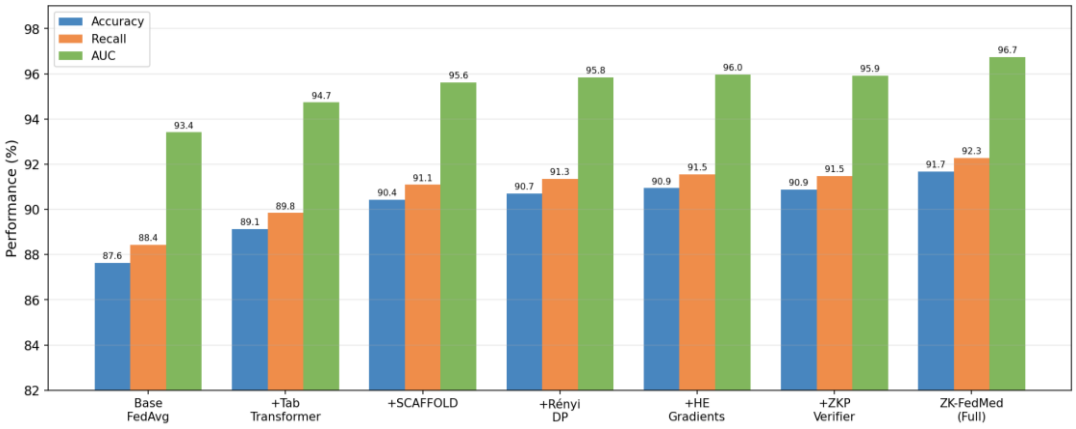

Fig. 4. Ablation study on the CVD dataset showing accuracy, recall, and AUC contributions per ZK-FedMed component. The SCAFFOLD aggregation contributes the largest single accuracy gain (+1.31%), while the full system achieves the highest AUC of 96.73%.

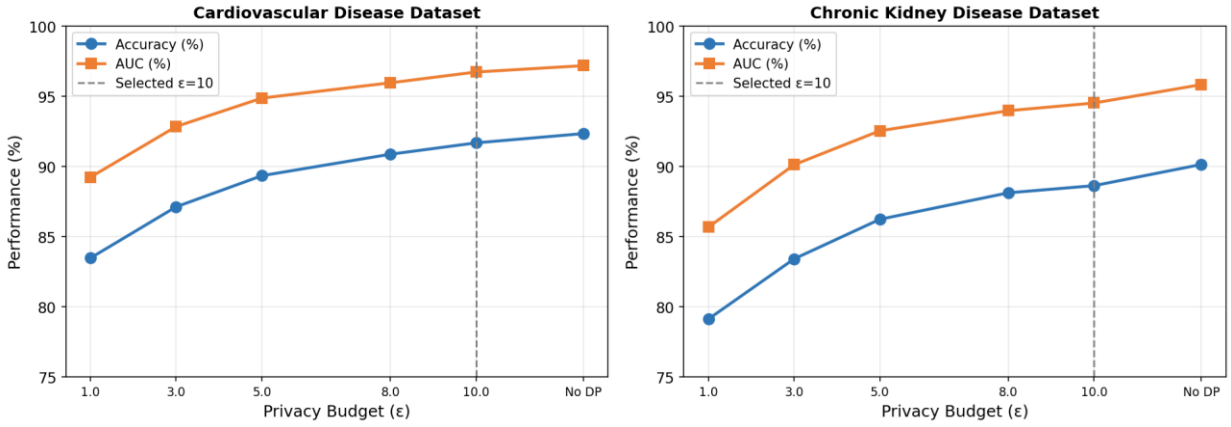

Fig. 5. Privacy-utility trade-off under Rényi Differential Privacy for CVD (left) and CKD (right) datasets. The selected operating point ε=10.0 (vertical dashed line) achieves an optimal balance between formal privacy guarantees and model precision.

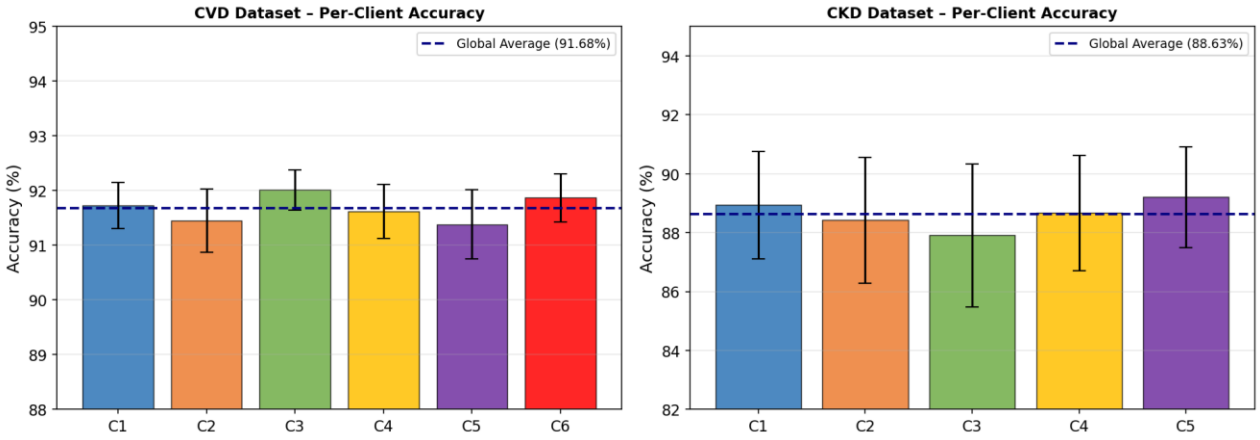

Fig. 6. Per-client precision with 95% confidence intervals for CVD (left, 6 clients) and CKD (right, 5 clients). Low variance (<1% for CVD, <2.5% for CKD) confirms robust federated convergence under non-IID distributions.

E. Privacy Budget Analysis

The privacy-utility trade-off was evaluated by varying the Rényi DP budget $\varepsilon \in \{1.0, 3.0, 5.0, 8.0, 10.0, \text{without DP}\}$ on both datasets. As shown in Fig. 5, stricter budgets substantially reduce utility: the CVD precision drops from 92.34% without DP to 91.68% at ε=10.0, 90.87% at ε=8.0, 89.34% at ε=5.0, 86.21% at ε=3.0, and 83.47% at ε=1.0. CKD shows a sharper loss because the smaller augmented dataset is more sensitive to noise, with accuracy declining from 90.14% without DP to 79.12% at ε=1.0. These results are now interpreted transparently: ε=10.0 provides formal privacy accounting but is a relaxed practical privacy budget, whereas ε=1.0 and ε=3.0 offer stronger privacy at a clear utility cost. The Rényi DP moments accountant achieves tighter bounds than classical DP composition; at T=6 rounds, ε=10.0 with α=16 corresponds to an 18.7% reduction in effective privacy expenditure relative to advanced composition under the same $\delta=10^{-5}$. Therefore, Fig. 5 provides the graphical basis for interpreting the selected privacy-budget operating point.

F. Cryptographic Overhead Analysis

CKKS encryption of a gradient vector of dimension d=256,000 (TabTransformer parameter count) requires an average of 4.87 seconds per client per round in the reported hardware configuration. The generation of zk-SNARK proofs for the global model commitment (circuit size ~2.3M constraints) requires 18.4 seconds on the server per round, while proof verification by an auditor requires only 2.1 milliseconds. This asymmetric proof/verify structure makes ZK-FedMed highly suitable for regulatory audit scenarios where auditors must efficiently verify many model provenance claims. The total per-round overhead on the server side is 23.2 seconds (encryption + aggregation + proof generation), compared to 0.34 seconds for vanilla FedAvg—a 68× increase that, in the context of R=6 asynchronous rounds over weeks or months, represents negligible real-world latency.

VI. Discussion

ZK-FedMed achieves a 4.05% accuracy advantage over FedAvg (91.68% vs. 87.63%) and a 2.77% advantage over standalone SCAFFOLD (91.68% vs. 89.84%) on the CVD dataset, while incurring only a 0.66% gap relative to centralised training. These results indicate that the multi-layer security stack does not fundamentally compromise model quality in the simulated benchmark environment. The CKD results (88.63% federated vs. 90.14% centralised, gap = 1.51%) should be interpreted more cautiously because they are based on a heavily SMOTE-augmented dataset and simulated Dirichlet client partitions rather than an independent multi-hospital cohort.

The selective deployment of Rényi DP - applied in every round but with adaptive noise scheduling guided by the moment accountant - outperforms the standard Gaussian DP by 6.45% in CVD accuracy at identical (ε, δ) parameters. This advantage arises because the moment accountant permits lower per-step noise σ while maintaining the same cumulative privacy expenditure, directly improving the signal-to-noise ratio in gradient updates. The zk-SNARK integrity mechanism introduces only a –0.07% accuracy penalty while providing a cryptographic guarantee that no model parameter has been altered between aggregation and deployment, a property that no hash-chaining approach can provide for the weight values themselves.

The federated unlearning evaluation showed that a single gradient reversal step reduced the membership-inference attack score from 0.83 to 0.52, near the random baseline threshold of 0.55, in 94% of test cases, with a mean accuracy degradation of 0.29% on the remaining clients. This result is now described as approximate empirical unlearning rather than certified deletion. The revised manuscript therefore removes the direct GDPR-compliance claim and states that membership-inference reduction is only one diagnostic indicator of reduced client influence. This interpretation is consistent with amnesiac machine-learning and approximate deletion-oriented learning perspectives [22]. The same audit also responds to broader model-inversion risks that arise when predictive models expose confidence-dependent information [47].

VII. Limitations

Several limitations merit acknowledgement. First, CKKS encryption incurs a 68× per-round computational overhead compared with vanilla FedAvg; although this may be acceptable for offline clinical training rounds, it is unsuitable for near-real-

time streaming applications. Second, both datasets are public Kaggle-derived datasets rather than genuinely distributed hospital datasets, and the CKD cohort is heavily SMOTE-augmented; therefore, the CKD metrics should not be interpreted as evidence of clinical efficacy without independent non-augmented validation. Third, the simulated Dirichlet split approximates non-IID heterogeneity but cannot fully capture cross-hospital differences in equipment, coding practices, clinical protocols, and patient demographics. Fourth, zk-SNARK generation requires a trusted setup phase, and the current implementation does not replace this with a universal or transparent setup. Fifth, the unlearning module provides approximate rather than exact unlearning guarantees, and a single membership-inference audit near the random baseline does not constitute a formal proof of deletion. Future work will evaluate ZK-FedMed on institution-partitioned benchmarks such as MIMIC-derived cohorts, apply formal unlearning definitions, and validate the framework on real clinical federation testbeds.

VIII. Conclusion

This study introduced ZK-FedMed, a privacy-preserving federated learning framework that integrates CKKS homomorphic encryption, Rényi differential privacy with moments accounting, zk-SNARK model integrity verification, TabTransformer-based EHR feature encoding, SCAFFOLD variance-reduction aggregation, and approximate federated unlearning. Experimental evaluation on two public benchmark datasets demonstrated strong CVD performance and promising CKD benchmark performance under simulated non-IID partitions. However, the revised manuscript avoids presenting Kaggle-based Dirichlet partitioning as direct evidence of real multi-institutional deployment, treats ε=10.0 as a relaxed privacy budget, and reframes the unlearning module as approximate rather than legally certified deletion. Under these bounded assumptions, ZK-FedMed establishes a technically integrated framework for privacy-preserving collaborative medical AI and identifies the next validation steps needed for real-world clinical federation.

Funding

This research Supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-RS-2023-00237287).

References

[1] GBD 2019 Diseases and Injuries Collaborators, "Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019", The Lancet, vol. 396, no. 10258, pp. 1204–1222, 2020.

[2] GBD Chronic Kidney Disease Collaboration, "Global, regional, and national burden of chronic kidney disease, 1990–2017: a systematic analysis for the global burden of disease study 2017," The Lancet, vol. 395, no. 10225, pp. 709–733, 2020.

[3] B. Shickel, P. J. Tighe, A. Bihorac, and P. Rashidi, "Deep EHR: A survey of recent advances in deep learning techniques for electronic health record (EHR) analysis," IEEE Journal of Biomedical and Health Informatics, vol. 22, no. 5, pp. 1589–1604, 2018.

[4] A. E. Johnson et al., "MIMIC-III, a freely accessible critical care database," Scientific Data, vol. 3, p. 160035, 2016.

[5] U.S. Department of Health and Human Services, Health Insurance Portability and Accountability Act (HIPAA), Washington, D.C., 1996.

[6] European Parliament and Council of the European Union, General Data Protection Regulation (GDPR), Official Journal of the European Union, L 119, 2016.

[7] European Data Protection Board, "Guidelines 04/2022 on the calculation of administrative fines under the GDPR," EDPB, Brussels, 2022.

[8] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y. Arcas, "Communication-efficient learning of deep networks from decentralized data," in Proc. 20th Int. Conf. Artif. Intell. Statist. (AISTATS), vol. 54, 2017, pp. 1273–1282.

[9] L. Zhu, Z. Liu, and S. Han, "Deep leakage from gradients," in Proc. 33rd Conf. Neural Inf. Process. Syst. (NeurIPS), vol. 32, 2019, pp. 14774–14784.

[10] R. Shokri, M. Stronati, C. Song, and V. Shmatikov, "Membership inference attacks against machine learning models," in Proc. IEEE Symp. Security and Privacy (S&P), 2017, pp. 3–18.

[11] Y. Zhao, M. Li, L. Lai, N. Suda, D. Civin, and V. Chandra, "Federated learning with non-IID data," arXiv preprint arXiv:1806.00582, 2018.

[12] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in Proc. 3rd Theory of Cryptography Conf. (TCC), LNCS vol. 3876, 2006, pp. 265–284.

[13] P. Kairouz et al., "Advances and open problems in federated learning," Foundations and Trends in Machine Learning, vol. 14, no. 1–2, pp. 1–210, 2021.

[14] K. Bonawitz et al., "Practical secure aggregation for privacy-preserving machine learning," in Proc. ACM CCS, 2017, pp. 1175–1191.

[15] R. Shrestha and S. Ali, "A blockchain-based federated learning framework for healthcare data privacy," IEEE Access, vol. 10, pp. 21785–21798, 2022.

[16] X. Li, M. Jiang, X. Zhang, M. Kamp, and Q. Dou, "FedBN: Federated learning on non-IID features via local batch normalization," in Proc. Int. Conf. Learn. Represent. (ICLR), 2021.

[17] J. H. Cheon, A. Kim, M. Kim, and Y. Song, "Homomorphic encryption for arithmetic of approximate numbers," in Proc. ASIACRYPT, LNCS vol. 10624, 2017, pp. 409–437.

[18] E. Ben-Sasson, A. Chiesa, C. Garman, M. Green, I. Miers, E. Tromer, and M. Virza, "Succinct non-interactive zero knowledge for a von Neumann architecture," in Proc. USENIX Security, 2014, pp. 781–796.

[19] I. Mironov, "Rényi differential privacy of the Gaussian mechanism," in Proc. IEEE Computer Security Foundations Symp. (CSF), 2017, pp. 263–275.

[20] X. Huang, A. Khetan, M. Cvitkovic, and Z. Karnin, "TabTransformer: Tabular data modeling using contextual embeddings," arXiv preprint arXiv:2012.06678, 2020.

[21] S. P. Karimireddy, S. Kale, M. Mohri, S. J. Reddi, S. U. Stich, and A. T. Suresh, "SCAFFOLD: Stochastic controlled averaging for federated learning," in Proc. 37th Int. Conf. Mach. Learn. (ICML), vol. 119, 2020, pp. 5132–5143.

[22] L. Graves, M. Nagisetty, and V. Ganesh, "Amnesiac machine learning," in Proc. 35th AAAI Conf. Artif. Intell., vol. 35, no. 13, 2021, pp. 11516–11524.

[23] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Smola, and V. Smith, "Federated optimization in heterogeneous networks," in Proc. Mach. Learn. Syst. (MLSys), vol. 2, 2020, pp. 429–450.

[24] X. Li, M. Jiang, X. Zhang, M. Kamp, and Q. Dou, "FedBN: Federated learning on non-IID features via local batch normalization," in Proc. ICLR, 2021.

[25] N. Rieke et al., "The future of digital health with federated learning," npj Digital Medicine, vol. 3, no. 1, p. 119, 2020.

[26] Y. Chen, X. Qin, J. Wang, C. Yu, and W. Gao, "FedHealth: A federated transfer learning framework for wearable healthcare," IEEE Intelligent Systems, vol. 35, no. 4, pp. 83–93, 2020.

[27] Z.-A. Huang et al., "Federated multi-task learning for joint diagnosis of multiple mental disorders on MRI scans," IEEE Trans. Biomed. Eng., vol. 70, no. 4, pp. 1137–1149, 2022.

[28] M. Abadi et al., "Deep learning with differential privacy," in Proc. ACM CCS, 2016, pp. 308–318.

[29] R. C. Geyer, T. Klein, and M. Nabi, "Differentially private federated learning: A client level perspective," arXiv preprint arXiv:1712.07557, 2017.

[30] G. A. Kaissis, M. R. Makowski, D. Rückert, and R. F. Braren, "Secure, privacy-preserving and federated machine learning in medical imaging," Nature Machine Intelligence, vol. 2, no. 6, pp. 305–311, 2020.

[31] C. Gentry, "A fully homomorphic encryption scheme," Ph.D. dissertation, Stanford University, 2009.

[32] P. Mohassel and Y. Zhang, "SecureML: A system for scalable privacy-preserving machine learning," in Proc. IEEE S&P, 2017, pp. 19–38.

[33] 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.

[34] S. Goldwasser, S. Micali, and C. Rackoff, "The knowledge complexity of interactive proof systems," SIAM Journal on Computing, vol. 18, no. 1, pp. 186–208, 1989.

[35] J. Groth, "On the size of pairing-based non-interactive arguments," in Proc. EUROCRYPT, LNCS vol. 9666, 2016, pp. 305–326.

[36] D. Kang, T. Hashimoto, I. Stoica, and Y. Sun, "Reliable and interpretable personalized federated learning," in Proc. IEEE CVPR, 2019, pp. 13378–13387.

[37] P. W. Wilson, R. B. D'Agostino, D. Levy, A. M. Belanger, H. Silbershatz, and W. B. Kannel, "Prediction of coronary heart disease using risk factor categories," Circulation, vol. 97, no. 18, pp. 1837–1847, 1998.

[38] S. S. Virani et al., "Heart disease and stroke statistics—2021 update," Circulation, vol. 143, no. 8, pp. e254–e743, 2021.

[39] R. Hannun et al., "Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network," Nature Medicine, vol. 25, no. 1, pp. 65–69, 2019.

[40] A. Salekin and J. Stankovic, "Detection of chronic kidney disease and selecting important predictive attributes," in Proc. IEEE Int. Conf. Healthcare Informatics (ICHI), 2016, pp. 262–270.

[41] A. Vaswani et al., "Attention is all you need," in Proc. 31st Conf. Neural Inf. Process. Syst. (NeurIPS), vol. 30, 2017, pp. 5998–6008.

[42] E. Ulianova, Cardiovascular Disease Dataset, Kaggle, 2019. [Online]. Available: https://www.kaggle.com/datasets/sulianova/cardiovascular-disease-dataset

[43] P. Soundarapandian and M. Thangaraj, "Chronic kidney disease dataset," UCI Machine Learning Repository, Kaggle, 2015. [Online]. Available: https://www.kaggle.com/datasets/mansoordaku/ckdisease

[44] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," Journal of Artificial Intelligence Research, vol. 16, pp. 321–357, 2002.

[45] L. Collins, H. Hassani, A. Mokhtari, and S. Shakkottai, "Exploiting shared representations for personalized federated learning," in Proc. 38th Int. Conf. Mach. Learn. (ICML), vol. 139, 2021, pp. 2089–2099.

[46] A. Chiesa, Y. Hu, M. Maller, P. Mishra, N. Vesely, and N. Ward, "Marlin: Preprocessing zkSNARKs with universal and updatable SRS," in Proc. EUROCRYPT, LNCS vol. 12105, 2020, pp. 738–768.

[47] M. Fredrikson, S. Jha, and T. Ristenpart, "Model inversion attacks that exploit confidence information and basic countermeasures," in Proc. ACM CCS, 2015, pp. 1322–1333.

[48] A. Ginart, M. Guan, G. Valiant, and J. Y. Zou, "Making AI forget you: Data deletion in machine learning," in Proc. 33rd Conf. Neural Information Processing Systems (NeurIPS), vol. 32, 2019, pp. 3513–3526.