Natural Resources for Human Health
eISSN: 2583-1194

Original Research

Received 24 July 2026
Revised 30 Aug 2026
Accepted 18 Sep 2026

Natural Resources for Human Health 2026, Vol. 6 (21s): 796–811

KEYWORDS:
Federated learning; Differential privacy; Post quantum cryptography; Secure aggregation; Content based image retrieval; Linear discriminant analysis

Quantum-Resilient Federated Learning with Differential Privacy for Secure Content-Based Image Retrieval in Cloud Environments

A. Neela Sundari[1]*, S. Sathiya[2], Gogineni Rajesh Chandra[3]

[1,2]Department of Computer Science and Engineering, Annamalai University, Chidambaram, Tamil Nadu, India
[3]Department of Computer Science and Engineering, NRI Institute of Technology, Guntur, Andhra Pradesh, India
*Corresponding author: A. Neela Sundari, ngargeya@gmail.com

ABSTRACT: Content based image retrieval (CBIR) in the cloud needs a model that describes images well, but the images often cannot be pooled, federated model updates can leak them, and the public key cryptography protecting those updates will not survive a quantum computer. This paper presents a retrieval framework that meets all three concerns in one round. Each client passes its images through a frozen public ResNet-50 and releases four summary statistics of the features with Gaussian noise for record level differential privacy. Secure aggregation with masks agreed through ML-KEM-768 sums the statistics, and the server builds a regularised linear discriminant projection from the noisy sum. On Caltech-101, CIFAR-10 and Tiny ImageNet the non-private version raises mean average precision from 79.5, 45.2 and 24.7 percent to 90.4, 78.8 and 43.4 percent, respectively. With a privacy budget of epsilon equal to 8 the cost is 7.2, 2.0 and 1.7 points, and it falls as the data per class grows. The result does not depend on how the data is split between clients, a membership inference attack stays near chance, and each client uploads 1.1 to 1.9 MB once, 91 to 93 percent less than 30 rounds of federated averaging. A single private round is therefore enough to build a useful shared retrieval space without pooling the images or exposing the model updates to a future quantum attacker

1. INTRODUCTION

Hospitals, research groups and companies keep very large image collections, and content based image retrieval, in short CBIR, lets a user find images by what they show rather than by a typed label. Deep networks have made CBIR much stronger because they turn every image into a feature vector in which similar images lie close together [1, 2]. The usual way to build such a service is to gather the images of every owner into one repository, train or tune a model there, and answer queries from the same place. That is exactly the step that rules, contracts and risk officers often forbid, because the images carry personal, clinical or commercial information [3, 4].

Federated learning keeps the images where they are. Each client trains on its own data and sends only a model update to a server that combines them [5]. The update is not harmless, however. A server that sees plain gradients can rebuild a good approximation of the training images [6], and a finished model can reveal whether a given image was used to train it [7]. Differential privacy, in short DP, puts a mathematical limit on this leakage by adding calibrated noise [8], and its best known form for deep learning, DP-SGD, clips and perturbs every per example gradient [9]. A third concern is the channel. The updates travel under keys agreed with RSA or elliptic curve cryptography, which a large quantum computer can break, so traffic recorded today can be opened later. NIST has now standardised ML-KEM, the key encapsulation mechanism that grew out of CRYSTALS-Kyber, as the replacement [10, 11].

These three protections are usually studied one at a time, and joining them in the usual way is expensive. DP-SGD needs many rounds of noisy training, every round is one more release that uses up privacy budget, and the noise hurts most when each

client holds only a little data. This paper takes a different route. The feature extractor is a public, frozen network, so it never needs private training. What has to be learnt from the private data is only the geometry of the retrieval space, and that geometry can be computed from four sums of the features. Sums are ideal for privacy engineering: each image changes them by a bounded amount, so the Gaussian mechanism applies directly; they can be added under secure aggregation, so the server never sees one client's contribution; and one round of communication is enough.

The contributions of this paper are as follows.

• A one round federated retrieval framework in which clients release noisy sufficient statistics of frozen features, and the server builds a regularised linear discriminant projection for retrieval. The result does not depend on how the data is split between clients.

• A record level differential privacy guarantee for the released statistics, with a sensitivity bound of 2 for L2 normalised features, calibrated with the analytic Gaussian mechanism and shared across clients through distributed noise.

• Secure aggregation in the ring of 64 bit integers with pairwise masks whose seeds are agreed with ML-KEM-768, and uploads protected by AES-256-GCM under ML-KEM-768 encapsulation, measured end to end on real statistics.

• An evaluation on three public datasets against a pooled data baseline, federated averaging and DP-FedAvg, including the privacy utility trade off, a strongly non-identical split, the communication and cryptographic cost, and an empirical membership inference test.

• Studies of the design choices that matter in practice: the amount of data per class, the number of clients and the role of secure aggregation in the noise, the dimension of the random projection, and the rule that sets the ridge.

The rest of the paper is arranged as follows. Section 2 reviews related work. Section 3 describes the framework. Section 4 gives the experimental setup, Section 5 the results, and Section 6 concludes.

2. RELATED WORK

2.1 Federated learning and differential privacy

Federated averaging trains a shared model by averaging models that clients have updated on their own data [5]. Differential privacy gives the standard definition of privacy for such systems [8], and DP-SGD applies it to deep learning by clipping every per example gradient and adding Gaussian noise, while tracking the privacy spent over many steps [9]. Phong et al. showed that plain gradients shared with a server leak the training data and protected them with additively homomorphic encryption [6]. Shokri et al. showed that a trained model itself reveals membership of training records [7]. The common weakness of these methods, in the retrieval setting studied here, is cost: many rounds, much communication, and noise that grows with the number of released updates.

2.2 Secure aggregation, secure computation and post quantum cryptography

Secure aggregation lets a server learn the sum of client vectors without seeing any single vector, using pairwise masks that cancel in the sum [12]. Secure multi party computation more generally allows parties to compute on data that none of them sees in full [13], and partial encryption of model updates can cut its cost [14]. For the quantum era, NIST standardised CRYSTALS-Kyber as the module lattice based key encapsulation mechanism ML-KEM [10]; its attack surface and mitigations have been studied in detail [11], and PQSF brought post quantum protection to federated learning [15]. Hamza combined post quantum key exchange with homomorphic processing of cloud images [16], and a recent survey reviews quantum resistant techniques for content based image retrieval in cloud computing environments [17].

2.3 Secure image retrieval in the cloud

SMedIR indexes encrypted medical images with a ConvNeXt backbone and searchable encryption [3]. Tian et al. added access control to encrypted retrieval with a tree index [4], Tang et al. supported user revocation in a mobile cloud [18], Wang et al. built similarity retrieval on homomorphic encryption with two servers [19], Chen et al. made query cost independent of the dataset size for cloud assisted IoT [20], and Jia et al. reduced the cost of homomorphic image classification by chunking [21]. These systems protect stored data and queries, but they train or tune their feature model on gathered data.

2.4 Gap

Table 1 places this work against representative earlier work. The retrieval systems gather data for training, the federated systems protect the channel with classical cryptography or protect the update without a formal guarantee, and the post quantum federated systems stop at classification. None of them learns a retrieval space from private data in a single round with a formal privacy guarantee and a quantum safe channel.

Table 1. Position of this work against representative earlier work.

Work	Data stays local	Formal DP	Quantum safe channel	Retrieval, not only classification	Rounds
FedAvg [5]	yes	no	no	no	many
DP-SGD [9]	no	yes	no	no	many
Phong et al. [6]	yes	no	no	no	many
PQSF [15]	yes	partly	yes	no	many
SMedIR [3]	no	no	no	yes	not applicable
Hamza [16]	no	no	yes	yes	not applicable
This work	yes	yes	yes	yes	one

3. PROPOSED FRAMEWORK

3.1 System and threat model

There are N clients, each holding a private labelled image collection, one aggregation server, and authorised users who send queries. The server is honest but curious: it follows the protocol and reads everything it receives. An outside adversary records all traffic, and may later own a large quantum computer, so any key agreed with RSA or elliptic curve cryptography must be treated as eventually broken. The aim is to learn a retrieval space from all collections together while every image stays with its owner and the released information satisfies record level differential privacy. **Figure 1** shows the framework, and Table 2 lists the symbols used in the rest of this section, from the clients, classes and feature dimension to the noise and the final projection.

Table 2. Notation used in the paper.

Symbol	Meaning
N, K, d	number of clients, number of classes, feature dimension ($d = 512$)
f, x	ResNet-50 feature (2048 values) and the projected, L2 normalised feature
R	fixed random orthogonal 2048×512 matrix from a public seed
s, S	sum of the features and sum of their outer products
c_k, n_k	feature sum and image count of class k
ε, δ	privacy budget and failure probability ($\delta = 10^{-5}$)
Δ_2, σ	L2 sensitivity of the released tuple and noise standard deviation of the sum
Σ_w, λ	within class scatter and ridge added before whitening
W, μ	retrieval projection and overall feature mean
L	number of values each client releases
k_{ij}	secret shared by clients i and j through ML-KEM-768

Natural Resources for Human Health
eISSN: 2583-1194

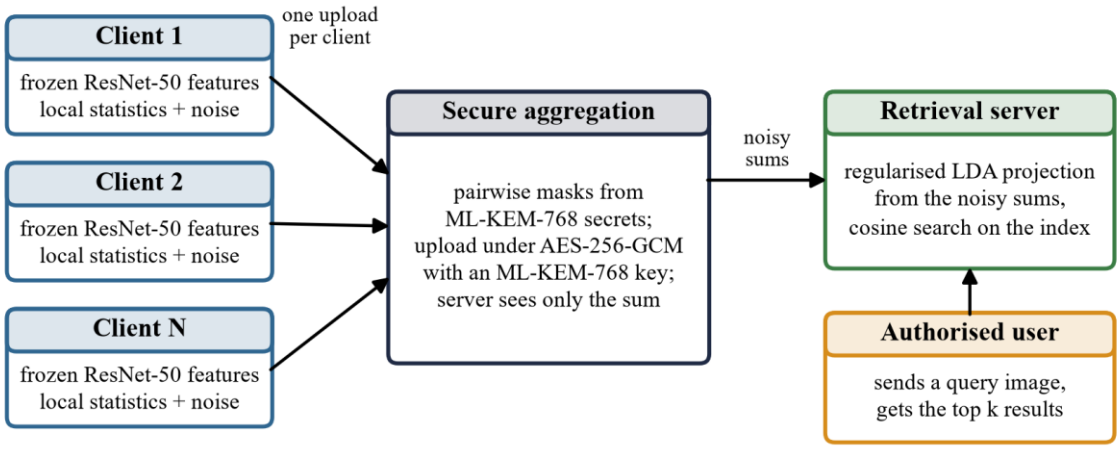


Figure 1. The proposed framework. Clients release noisy sums of frozen features once, the sums are added under secure aggregation, and the server builds the retrieval projection.

3.2 Shared feature extractor

Every client uses the same ResNet-50 with public ImageNet weights, frozen and never trained on private data [1]. Each image is resized to 224×224 and the 2048 dimensional pooled feature is taken. The feature is L2 normalised, multiplied by a fixed random orthogonal 2048×512 matrix R that is generated from a public seed and depends on no data, and L2 normalised again:

$$x = \frac{R^\top \hat{f}}{\|R^\top \hat{f}\|_2}, \quad \hat{f} = \frac{f}{\|f\|_2}. \tag{1}$$

The projection shrinks the released statistics sixteen fold and the noise they need, and because $\|x\|_2 = 1$ every image has bounded influence.

3.3 Federated sufficient statistics

Client i, holding pairs (x, y) with $y \in \{1, \dots, K\}$, computes four sums: the feature sum s_i, the second moment S_i, the per class sums $c_{i,k}$ and the per class counts $n_{i,k}$:

$$s_i = \sum x, \;\; S_i = \sum x x^\top, \;\; c_{i,k} = \sum_{y=k} x, \;\; n_{i,k} = \sum_{y=k} 1. \tag{2}$$

Because all four are sums, the server needs only their totals over clients, $s = \sum_i s_i$ and so on, which is what secure aggregation delivers. The totals are the same whatever way the data is split between clients, so the method is untouched by non identically distributed data.

3.4 Differential privacy

Adding or removing one image changes s by x, S by $xx^\top$, one row of c by x and one count by 1. With $\|x\|_2 = 1$ each change has L2 norm at most 1, since $\|xx^\top\|_F = \|x\|_2^2$, so the whole released tuple has L2 sensitivity

$$\Delta_2 \le \sqrt{1 + 1 + 1 + 1} = 2. \tag{3}$$

Releasing the tuple with Gaussian noise of standard deviation σ on every entry is (ε, δ) differentially private when σ satisfies the analytic Gaussian mechanism condition [22]

$$\Phi\left(\frac{\Delta_2}{2\sigma} - \frac{\varepsilon\sigma}{\Delta_2}\right) - e^{\varepsilon}\,\Phi\left(-\frac{\Delta_2}{2\sigma} - \frac{\varepsilon\sigma}{\Delta_2}\right) \le \delta, \tag{4}$$

and σ is the smallest such value. Only the upper triangle of S is released, and the noise matrix is mirrored to keep S symmetric. Each client adds noise of standard deviation $\sigma/\sqrt{N}$, so the aggregated sum carries exactly $\mathcal{N}(0, \sigma^2)$ per entry and the guarantee

holds for the sum that the server sees. This relies on secure aggregation, which is why the two are designed together. The statistics are released once, so there is no composition across rounds.

3.5 Secure aggregation over a post quantum channel

Every client holds an ML-KEM-768 key pair [10]. For each pair of clients $i < j$, client i encapsulates to the public key of j, the server relays the 1,088 byte capsule, and both obtain a shared secret k_{ij}. The released vector u_i is encoded in fixed point with scale 2^{24} as a vector of L integers modulo 2^{64} and masked as

$$m_i = u_i + \sum_{j>i} \mathrm{PRG}\,(k_{ij}) - \sum_{j<i} \mathrm{PRG}\,(k_{ji}) \bmod 2^{64}, \tag{5}$$

where PRG is the ChaCha20 key stream [12]. The masked vector is uploaded under AES-256-GCM with a key obtained by encapsulating to the server's ML-KEM-768 public key. The server decapsulates, decrypts and adds, the masks cancel, and it decodes the exact sum without seeing any single u_i. Dropped clients are handled by the standard recovery step of secure aggregation [12], which is not exercised in the experiments.

3.6 Retrieval projection

From the noisy totals the server forms the class means $M_k = c_k/n_k$, the overall mean $\mu = s/n$ and the within class scatter

$$\Sigma_w = \frac{1}{n}(S - \sum_k n_k\, M_k M_k^{\top}). \tag{6}$$

It whitens with $(\Sigma_w + \lambda I)^{-1/2}$ and keeps the leading directions of the whitened between class scatter, which is regularised linear discriminant analysis [23]. Noise inflates the spectrum of Σ_w, so the ridge is set from the noise level alone, never from test data: the spectral norm of a symmetric Gaussian matrix of size d with entry deviation σ/n is about $2\sigma\sqrt{d}/n$, giving

$$\lambda = 10^{-3} + \frac{2\sigma\sqrt{d}}{n}. \tag{7}$$

Up to $\min(K - 1{,}128)$ directions are kept. The resulting projection W is shared with every client and with the retrieval service.

3.7 Retrieval

The gallery is indexed once as $z = W^{\top}(x - \mu)/\|W^{\top}(x - \mu)\|_2$. A query image passes through the same extractor and projection, and gallery images are ranked by cosine similarity.

3.8 Protocol summary and cost

Table 3 lists the nine steps of one complete run. Steps 1 and 2 set up the keys and happen once for a group of clients; steps 3 to 6 run at each client and need no further message from the server; steps 7 and 8 run at the server; step 9 is the retrieval service itself. The table shows that a client sends exactly one upload in the whole protocol, apart from the small capsules of the key setup, and that the server never receives a value from which a single client's statistics could be read.

Table 3. Steps of one complete run of the proposed protocol.

Step	Party	Operation
1. Key setup	every client	generate an ML-KEM-768 key pair and publish the 1,184 byte public key
2. Pairwise secrets	each client pair	for $i < j$, client i encapsulates to client j; the server relays the 1,088 byte capsule and both obtain k_{ij}
3. Local statistics	every client	extract x for every image and compute s_i, S_i, $c_{i,k}$ and $n_{i,k}$
4. Noise	every client	add Gaussian noise of standard deviation $\sigma/\sqrt{N}$ to every released value
5. Masking	every client	encode in fixed point with scale 2^{24} modulo 2^{64} and add the pairwise ChaCha20 masks
6. Upload	every client	encrypt under AES-256-GCM with a key encapsulated to the server's ML-KEM-768 public key and send once
7. Aggregation	server	decapsulate, decrypt and add the uploads; the masks cancel and the noisy sum is decoded
8. Projection	server	form Σ_w, set λ from σ, compute W and share W and μ
9. Retrieval	service	index the gallery with W and rank gallery images for each query by cosine similarity

Each client releases the vector s_i with d values, the upper triangle of S_i, the class sums and the class counts, so the number of released values is

$$L = d + \frac{d(d+1)}{2} + Kd + K. \tag{8}$$

With $d = 512$ this gives 183,140 values for the 100 classes of Caltech-101, 136,970 for CIFAR-10 and 234,440 for the 200 classes of Tiny ImageNet, each stored as an 8 byte integer. A client computes its second moment in $O(n_i d^2)$ operations for n_i images, and the server needs two eigendecompositions of size d, $O(d^3)$, independent of the number of images and clients. The upload grows with d^2, which is why the random projection to 512 dimensions matters; Section 5 measures this choice.

4. EXPERIMENTAL SETUP

4.1 Datasets

Three public datasets are used (Table 4). CIFAR-10 has 60,000 small images in ten classes [24]. Caltech-101 is used without its background class and without the Faces_easy class, which holds tighter crops of the same photographs as the Faces class [25]; a fixed stratified 80/20 split with seed 0 gives 6,593 gallery and 1,649 query images. Tiny ImageNet has 200 classes [26], with 500 training and 50 validation images in each class; the validation set serves as the query set. In every case the training images form the gallery and the test images form the queries. Tiny ImageNet replaces the ImageNet-50 subset used in an earlier draft of this study, because ImageNet requires an access approval and a fifty class subset is not a standard, reproducible benchmark. Table 4 also shows how different the three sets are for a privacy method: with ten clients, a client holds about 6.6 images of each class on Caltech-101, 50 on Tiny ImageNet and 500 on CIFAR-10, and Caltech-101 classes range from 26 to 640 training images. These numbers explain much of the privacy cost reported in Section 5.

Table 4. Datasets, splits and the per class data available to the clients.

	Caltech-101	CIFAR-10	Tiny ImageNet
Classes	100	10	200
Gallery (train) images	6,593	50,000	100,000
Query (test) images	1,649	10,000	10,000
Train images per class	26 to 640	5,000	500
Per class and client (IID)	about 6.6	500	50

4.2 Clients and splits

Every experiment uses $N = 10$ clients. In the IID split the training set is shuffled and cut into ten equal parts. In the non-IID split each class is divided among the clients with Dirichlet proportions of concentration 0.1, so most clients hold only a few classes.

4.3 Compared methods

Six methods are compared. Frozen features ranks by cosine similarity on x with no learning. Central head trains a linear head ($512 \to 128$ with a cosine classifier) on all data pooled, for 30 epochs with SGD at learning rate 0.05; it is the non-private upper reference. FedAvg head trains the same head by federated averaging for 30 rounds of one local epoch [5]. DP-FedAvg head adds record level DP-SGD at every client [9], with Poisson sampling, clipping norm 1, batch size 256 and learning rate 1.0, and the noise multiplier set by the Rényi accountant for the chosen ε. The proposed method is run without noise and with $\varepsilon \in \{1,3,8\}$. Throughout, $\delta = 10^{-5}$, and every noisy or trained method is run with three seeds.

Table 5 gives the noise behind every private result. For the proposed method the noise depends only on ε, because the sensitivity is fixed at 2: the sum carries a standard deviation of 1.200, 2.781 and 7.461 at $\varepsilon = 8$, 3 and 1, and each of the ten clients adds $1/\sqrt{10}$ of it. The ridge λ then follows from Eq. (7) and grows as the data shrinks, so it is largest for Caltech-101. The DP-FedAvg noise multipliers are smaller numbers, but they are applied at every one of many steps, which is the cost the one round design avoids.

Table 5. Noise used at each privacy budget ($\delta = 10^{-5}$, 10 clients). σ is the standard deviation carried by the aggregated sum, $\sigma/\sqrt{N}$ the share added by each client, λ the ridge set from the noise, and the last three columns the noise multiplier of DP-FedAvg.

ε	σ (sum)	$\sigma/\sqrt{N}$	λ Caltech	λ CIFAR	λ Tiny	DP-FedAvg Caltech	DP-FedAvg CIFAR	DP-FedAvg Tiny
8	1.200	0.380	0.0092	0.0021	0.0015	2.52	1.09	0.88
3	2.781	0.879	0.0201	0.0035	0.0023	5.66	2.07	1.53
1	7.461	2.359	0.0521	0.0078	0.0044	15.16	5.23	3.67

4.4 Measures

Retrieval quality is measured by mean average precision, mAP, over the full ranking of the gallery, and by the precision in the first one, ten and hundred results, P@1, P@10 and P@100. Leakage is tested empirically with a loss threshold membership inference attack [7]: 5,000 training images (members) and 5,000 test images (non members) are scored by the loss of a cosine classifier with scale 16, using the class weights for the trained heads and the projected noisy class means for the proposed method, and the area under the ROC curve, AUC, is reported. Cost is measured by the bytes each client uploads and the time of each cryptographic step. All experiments ran on a Dell G15 laptop with an Intel Core i7-12700H, 8 GB of memory and an NVIDIA RTX 3050 laptop GPU with 4 GB, under Windows 11 with Python 3.12, PyTorch 2.9, and the pqcrypto and cryptography libraries for ML-KEM-768, ChaCha20 and AES-256-GCM.

4.5 Additional studies

Four further studies use the same features and seeds. In the first, the statistics of CIFAR-10 are computed from only 50 to 5,000 training images per class while the gallery stays the full training set, so that only the projection changes. In the second, the number of clients is set to 5, 10, 20 and 50, and the distributed noise of the proposed design is compared with the case without secure aggregation, where every client must add the full σ itself. In the third, the dimension of the random projection is varied from 128 to 2048, the last meaning no projection. In the fourth, the ridge rule of Eq. (7) is compared with fixed values of λ.

Natural Resources for Human Health
eISSN: 2583-1194

5. RESULTS AND DISCUSSION

5.1 Retrieval quality

Table 6 gives the main results. Learning a retrieval space from the private data matters: the proposed method without noise lifts mAP over the frozen features by 10.9, 33.5 and 18.7 points on Caltech-101, CIFAR-10 and Tiny ImageNet. A single round of statistics is enough to pass the head trained on pooled data on Caltech-101 and CIFAR-10 and to come within 1.1 points of it on Tiny ImageNet (90.2, 75.3 and 44.6 percent for the pooled head) and to beat federated averaging after 30 rounds (85.8, 74.2 and 43.2 percent). The reason is that, for a frozen extractor, the class conditional first and second moments already carry most of the information a linear retrieval space needs, and computing them exactly avoids the optimisation error of federated training. With 200 classes and 500 images per class, Tiny ImageNet is where a trained head gains most from its extra flexibility.

Table 6. Retrieval quality on the three test sets (IID split over 10 clients). mAP is the mean of three runs with the standard deviation; P@10 is the precision in the first ten results. All values in percent.

Method	Caltech mAP	Caltech P@10	CIFAR mAP	CIFAR P@10	Tiny mAP	Tiny P@10
Frozen features, no learning	79.5	89.1	45.2	81.1	24.7	55.2
Central head, all data pooled	90.2 $\pm$ 0.3	93.1	75.3 $\pm$ 0.3	86.9	44.6	67.3
FedAvg head, 30 rounds	85.8 $\pm$ 0.2	91.2	74.2 $\pm$ 0.5	85.8	43.2	66.0
Proposed, one round, no DP	90.4	93.5	78.8	84.1	43.4	64.7
Proposed, DP, $\varepsilon = 8$	83.2 $\pm$ 0.1	89.6	76.8	83.2	41.7	63.9
Proposed, DP, $\varepsilon = 3$	79.0 $\pm$ 0.2	87.7	74.8 $\pm$ 0.1	82.3	39.0	62.4
Proposed, DP, $\varepsilon = 1$	75.4 $\pm$ 0.2	86.3	71.3 $\pm$ 0.1	80.9	32.1	58.3
DP-FedAvg head, $\varepsilon = 8$	77.4 $\pm$ 0.3	87.3	71.9 $\pm$ 0.2	86.3	34.4 $\pm$ 0.1	60.3
DP-FedAvg head, $\varepsilon = 3$	75.1 $\pm$ 0.3	86.4	66.1 $\pm$ 0.4	84.8	29.7 $\pm$ 0.2	57.1
DP-FedAvg head, $\varepsilon = 1$	73.7 $\pm$ 0.4	85.9	56.9 $\pm$ 0.5	82.1	23.3 $\pm$ 0.2	52.3

Table 7 looks at the top of the ranking and deeper into it. At the first result the frozen features are already strong (92.7, 85.7 and 61.2 percent), because the nearest neighbour of a query is often a very similar image, and the methods differ by only a few points. On CIFAR-10 the proposed method is in fact the weakest at P@1, 84.1 percent against 85.7 for the frozen features and 87.3 for the central head. With ten classes the projection keeps only $K - 1 = 9$ directions, which places images by class very well but keeps little of the fine detail that decides the single nearest neighbour. Deeper in the ranking the learnt projection matters much more: at P@100 the frozen features give 60.8, 74.8 and 44.2 percent and the proposed method without DP gives 65.4, 84.0 and 59.4 percent, close to the central head. P@100 on Caltech-101 is low for every method because 94 of its 100 classes have fewer than 100 gallery images. The private version keeps most of the gain at $\varepsilon = 8$ and loses more at $\varepsilon = 1$, in the same order as mAP.

Natural Resources for Human Health
eISSN: 2583-1194

Table 7. Precision at the first result (P@1) and in the first hundred results (P@100), IID split, mean of the runs, in percent.

Method	Caltech P@1	Caltech P@100	CIFAR P@1	CIFAR P@100	Tiny P@1	Tiny P@100
Frozen features	92.7	60.8	85.7	74.8	61.2	44.2
Central head	94.5	65.3	87.3	85.8	69.7	62.0
FedAvg head	93.6	63.8	86.5	84.7	68.7	60.5
DP-FedAvg head, $\varepsilon = 8$	92.0	59.9	88.2	84.4	64.0	52.8
Proposed, no DP	94.8	65.4	84.1	84.0	67.8	59.4
Proposed, DP, $\varepsilon = 8$	92.6	62.8	83.2	83.0	67.0	58.1
Proposed, DP, $\varepsilon = 1$	90.7	59.0	81.4	80.4	62.7	50.0

5.2 Privacy and utility

Figure 2 shows mAP against the privacy budget. With $\varepsilon = 8$ the proposed method loses 7.2, 2.0 and 1.7 points against its own non-private version, and still stays well above the frozen features on CIFAR-10 and Tiny ImageNet. On Caltech-101 the loss is largest: the private version is 79.0 percent at $\varepsilon = 3$, about the level of the frozen features (79.5), and 75.4 at $\varepsilon = 1$. The ordering follows the amount of data per class: Caltech-101 offers about 65 training images per class, Tiny ImageNet 500 and CIFAR-10 5,000, while the noise needed for a given ε is fixed, so the class means of a small class are the first to be drowned. At $\varepsilon = 1$ the loss grows to 15.0, 7.5 and 11.3 points. DP-FedAvg, spending the same budget of $\varepsilon = 8$ over many noisy steps, reaches 77.4, 71.9 and 34.4 percent, below the proposed method on every dataset. The run to run standard deviation of the private version is at most 0.2 points, because the only randomness is the noise itself.

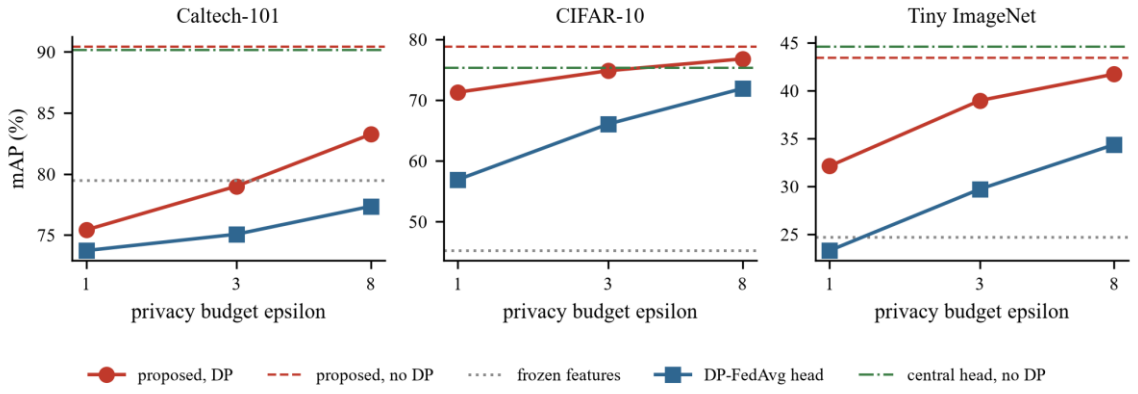


Figure 2. mAP against the privacy budget ε ($\delta = 10^{-5}$), with the non-private references.

5.3 Data per class

The explanation above can be tested directly. **Figure 3** computes the statistics of CIFAR-10 from 50 to 5,000 training images per class, with the full gallery kept for retrieval. Without noise, 50 images per class already give 72.4 percent mAP, against 78.8 with all 5,000. With noise the curve is much steeper: at $\varepsilon = 8$ the method gives 54.0, 69.2 and 76.8 percent with 50, 500 and 5,000 images per class, and at $\varepsilon = 1$ it gives 28.2, 59.4 and 71.3 percent. With 100 or fewer images per class the private version at $\varepsilon = 1$ falls under the frozen features (45.2 percent). The noise added for a given ε does not depend on the data, so its effect shrinks as the sums grow, and the amount of data per class, not the number of classes, decides whether a privacy budget is affordable.

Natural Resources for Human Health
eISSN: 2583-1194

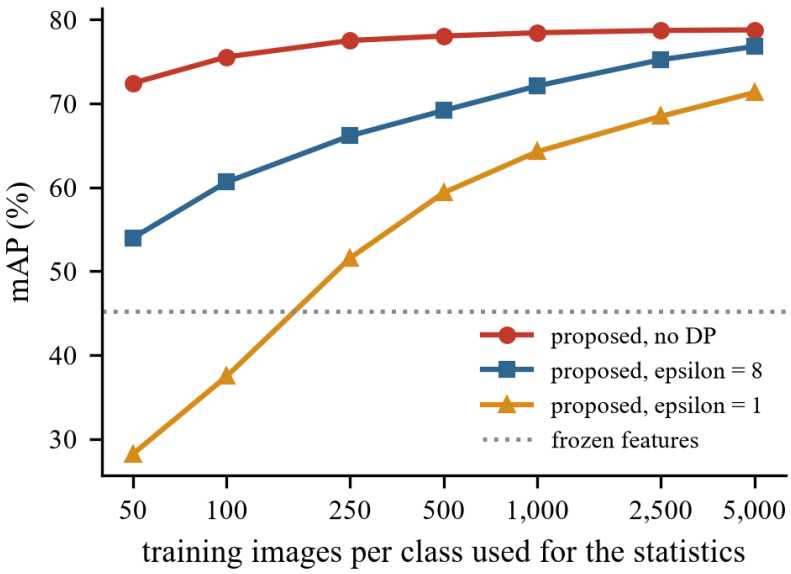


Figure 3. CIFAR-10 mAP when the statistics come from a given number of training images per class (gallery unchanged).

5.4 Number of clients and the role of secure aggregation

Figure 4 varies the number of clients at $\varepsilon = 8$. With secure aggregation each client adds only $\sigma/\sqrt{N}$, the sum always carries σ, and mAP stays at 83.3 to 83.3 percent on Caltech-101 from 5 to 50 clients. Without secure aggregation every client must make its own release private, so it adds the full σ and the sum carries $\sigma\sqrt{N}$. mAP then falls from 79.3 to 75.3 percent on Caltech-101, from 74.9 to 70.8 on CIFAR-10 and from 39.1 to 31.0 on Tiny ImageNet. Secure aggregation is therefore not only a confidentiality measure; it is what lets the privacy cost stay flat as more clients join.

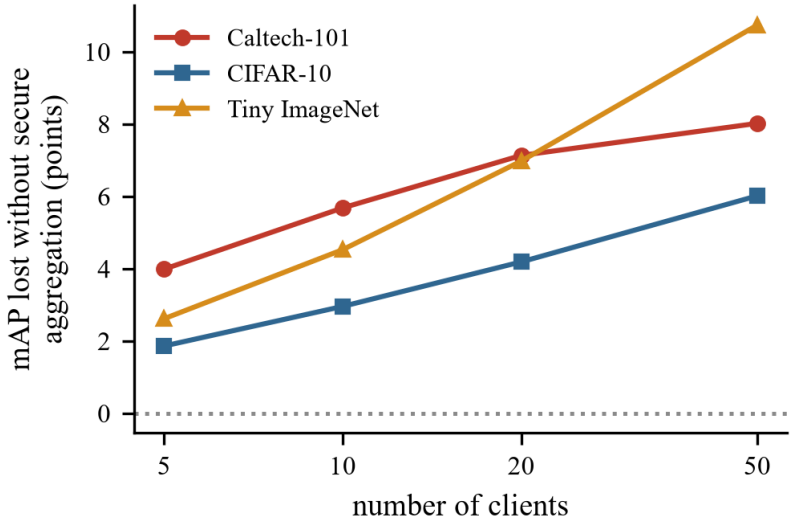


Figure 4. mAP lost at $\varepsilon = 8$ when every client adds the full noise itself, compared with distributed noise under secure aggregation, whose mAP stays flat, for 5 to 50 clients.

5.5 Membership inference

The formal guarantee is the main protection, but an empirical attack is a useful check. Table 8 gives the AUC of the loss threshold attack. Every method is close to chance: the trained heads reach at most 0.538, the proposed method without DP at

most 0.519, and the private versions between 0.504 and 0.508. The highest value of all belongs to the central head on Caltech-101, the smallest dataset. A frozen extractor followed by a linear map or a small head has little capacity to memorise single images, so this simple attack learns little even without noise. The test cannot prove privacy, and stronger attacks may do better; the guarantee in Section 3 holds whatever attack is used.

Table 8. Area under the ROC curve of a loss threshold membership inference attack (5,000 members against test images; 0.5 is chance).

Method	Caltech-101	CIFAR-10	Tiny ImageNet
Central head	0.538	0.512	0.526
FedAvg head	0.524	0.510	0.520
DP-FedAvg, $\varepsilon = 8$	0.509	0.509	0.510
DP-FedAvg, $\varepsilon = 1$	0.503	0.509	0.505
Proposed, no DP	0.519	0.505	0.503
Proposed, $\varepsilon = 8$	0.506	0.508	0.506
Proposed, $\varepsilon = 3$	0.504	0.507	0.506
Proposed, $\varepsilon = 1$	0.505	0.506	0.506

5.6 Non-identical data

Figure 5 compares the IID split with the Dirichlet split of concentration 0.1. Federated averaging loses 0.9, 2.3 and 2.6 points, as expected when clients hold few classes. The proposed method returns the same numbers to every digit, which is a check on the implementation rather than a coincidence: the aggregated sums do not depend on how the data is divided, and with the same number of clients the noise is drawn the same way. The same holds with noise: at $\varepsilon = 8$, 3 and 1 the private version gives the same mAP on both splits for every dataset and seed.

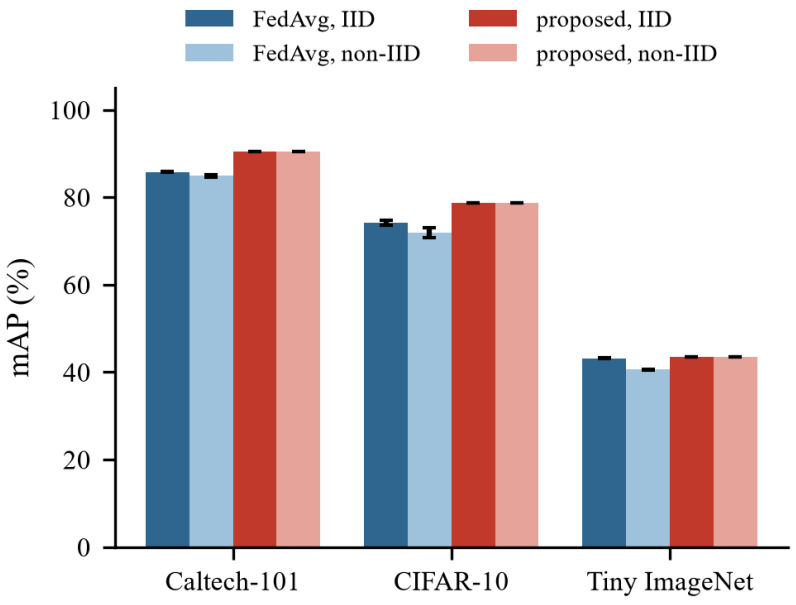


Figure 5. mAP with the IID split and the Dirichlet 0.1 split, without DP.

5.7 Communication and cryptographic cost

Table 9 and **Figure 6** report the cost of one complete aggregation. Each client uploads its masked statistics once, 1.1 to 1.9 MB depending on the number of classes, against 16 to 22 MB for the 30 rounds of federated averaging, a saving of 91 to 93 percent. The post quantum protection adds 1,116 bytes per upload, the ML-KEM-768 capsule plus the GCM tag and nonce, and the pairwise mask setup adds about 9.8 kB per client once. Masking takes 24 to 41 ms per client and the server aggregates ten uploads in 20 to 33 ms. The decoded sum differs from the plain sum by at most 2.2×10^{-7}, the fixed point rounding. Table 10 compares key establishment options: ML-KEM-768 costs 0.16 ms for encapsulation and decapsulation together, about the same as X25519 and faster than RSA-3072, while being the only single scheme that is quantum safe. Its extra size, about a kilobyte, is negligible next to the payload. The communication saving therefore comes from the one round design and not from the cipher: a post quantum channel costs slightly more bytes than a classical one, never fewer.

Table 9. Cost of one secure, post quantum protected aggregation (10 clients, $\varepsilon = 8$). Upload counts the masked statistics, the ML-KEM-768 capsule, the GCM tag and the nonce. The FedAvg figure counts the head sent down and up in each of 30 rounds.

Dataset	Values per client	Upload (MB)	FedAvg, 30 rounds (MB)	Saving (%)	Masking (ms)	Max decode error
Caltech-101	183,140	1.47	18.8	92	29	2.2×10^{-7}
CIFAR-10	136,970	1.10	16.1	93	24	2.2×10^{-7}
Tiny ImageNet	234,440	1.88	21.9	91	41	2.2×10^{-7}

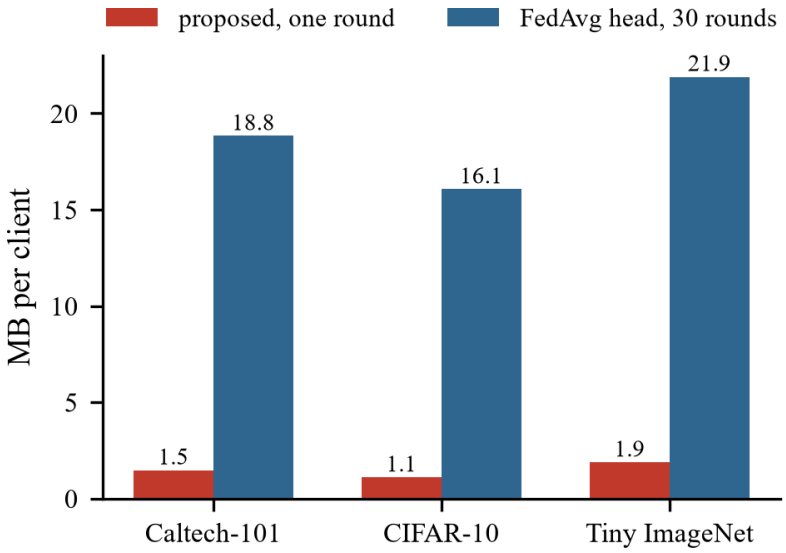


Figure 6. Upload per client: one round of the proposed method against 30 rounds of federated averaging.

Table 10. Key establishment options for one upload, measured on the test laptop (median of 30 runs). Bulk encryption is AES-256-GCM in every case.

Scheme	Public key (B)	Per upload (B)	Time (ms)	Quantum safe
RSA-3072 OAEP key transport	422	384	1.99	no
X25519 key agreement	32	32	0.16	no
ML-KEM-768 (proposed)	1,184	1,088	0.16	yes
X25519 + ML-KEM-768	1,216	1,120	0.31	yes

5.8 Design choices

Table 11 varies the dimension of the random projection. Without noise a larger d helps throughout: 2048 dimensions, that is no projection, give 92.3, 81.1 and 46.8 percent against 90.4, 78.8 and 43.4 at 512. With $\varepsilon = 8$ the picture changes, because the noise grows with the number of released values: 1024 dimensions are best on CIFAR-10 and Tiny ImageNet (77.1 and 43.3 percent), 512 is best on Caltech-101, and 2048 is never best. Small projections lose a lot: 128 dimensions cost 5.8 to 17.7 points against 512. The upload grows with d^2, from 1.10 MB at 512 to 4.29 MB at 1024 for CIFAR-10. The value of 512 used in the paper is therefore a balance between accuracy and upload size, not the most accurate choice; a deployment with ample bandwidth can use 1024 for a gain of up to 1.5 points at $\varepsilon = 8$.

Table 11. Effect of the dimension d of the fixed random projection on mAP (percent) without DP and with $\varepsilon = 8$, and the resulting upload for CIFAR-10.

d	Caltech no DP	Caltech $\varepsilon = 8$	CIFAR no DP	CIFAR $\varepsilon = 8$	Tiny no DP	Tiny $\varepsilon = 8$	CIFAR upload (MB)
128	80.0	77.4	69.3	68.8	25.8	25.7	0.08
256	87.2	81.8	75.5	74.4	36.9	36.1	0.29
512 (used)	90.4	83.2	78.8	76.8	43.4	41.7	1.10
1024	91.7	82.9	80.3	77.1	46.1	43.3	4.29
2048 (no projection)	92.3	81.9	81.1	77.0	46.8	42.8	16.97

Table 12 compares the noise rule for the ridge, Eq. (7), with fixed values. On Caltech-101 the rule is within 0.1 points of the best fixed value. On CIFAR-10 and Tiny ImageNet a smaller fixed ridge of 10^{-3} does better, by 1.8 and 1.2 points at $\varepsilon = 8$, while larger fixed values are clearly worse. The rule is therefore conservative: it protects against a badly conditioned scatter matrix when data is scarce, but it over regularises when every class has hundreds of images. The rule is kept in all other results because it depends only on the noise level and the noisy count, never on the queries; picking the best value from this table would mean tuning on the test queries.

Table 12. mAP (percent) with fixed ridge values and with the noise based rule, at $\varepsilon = 8$ / $\varepsilon = 1$.

Ridge λ	Caltech-101	CIFAR-10	Tiny ImageNet
fixed 10^{-3}	82.9 / 75.5	78.6 / 74.1	43.0 / 33.3
fixed 10^{-2}	83.2 / 75.5	71.5 / 70.6	35.3 / 30.4
fixed 10^{-1}	82.4 / 75.4	66.5 / 66.3	31.0 / 26.9
noise rule, Eq. (7)	83.2 / 75.4	76.8 / 71.3	41.7 / 32.1

5.9 Security analysis

Against a curious server, secure aggregation reveals only the sum, and each pairwise mask is a ChaCha20 key stream under a secret agreed by ML-KEM-768, which the server relays but cannot open. Against inference from what is released, the sum carries Gaussian noise that gives (ε, δ) record level differential privacy with the sensitivity bound derived in Section 3, and anything computed from it, including the projection and the index, inherits the guarantee as post processing [8]. Against an eavesdropper with a future quantum computer, every key in the system comes from ML-KEM-768, whose security rests on the module learning with errors problem [10, 11]; recorded traffic therefore stays protected. The feature extractor is public and frozen, so it carries no private information.

5.10 Limitations

The extractor is frozen, so the method cannot adapt the features themselves to a domain that is far from ImageNet; a medical collection, for instance, may need a public extractor pretrained on medical images. The projection is linear. The differential privacy guarantee covers the released statistics and relies on secure aggregation for the distributed noise; if a client could see the other clients' shares the noise would have to be added in full at every client, which costs a factor of $\sqrt{N}$ in noise. Client dropout and malicious clients are outside the experiments, and the gallery index itself is not encrypted. The guarantee covers one release: if the same clients later release new statistics, for example after collecting more images, the two releases compose and the total budget grows. Finally, the ridge rule and the projection size were fixed before the experiments and are not the best values in every case, as the design study shows.

6. CONCLUSION AND FUTURE WORK

This paper showed that a retrieval space can be learnt from private, distributed image collections in a single round, with a formal record level privacy guarantee and a channel that stays safe against a quantum adversary. Clients release noisy sums of frozen features, secure aggregation over ML-KEM-768 adds them without exposing any client, and the server builds a regularised linear discriminant projection. Without noise the method comes close to or passes training on pooled data and does better than federated averaging; with $\varepsilon = 8$ it costs 2.0 and 1.7 points on CIFAR-10 and Tiny ImageNet and more on the small classes of Caltech-101; and each client uploads about a megabyte once. The studies show that the privacy cost is decided by the amount of data per class, that secure aggregation keeps it flat as clients are added, and that an attack on membership stays close to chance. Future work will add dropout recovery and robustness to malicious clients, study public extractors pretrained on medical images, and extend the projection beyond the linear case.

ACKNOWLEDGEMENTS

The authors thank the providers of the Caltech-101, CIFAR-10 and Tiny ImageNet datasets for making them publicly available.

AUTHOR CONTRIBUTIONS

A. Neela Sundari: Conceptualization, Methodology, Software, Validation, Formal Analysis, Writing - Original Draft.

S. Sathiya: Supervision, Writing - Review and Editing.

Gogineni Rajesh Chandra: Conceptualization, Supervision, Project Administration, Writing - Review and Editing.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

ETHICS STATEMENT

This study used only publicly available image datasets. No new data were collected from humans or animals, so ethical approval was not required.

DATA AVAILABILITY STATEMENT

The Caltech-101 [25], CIFAR-10 [24] and Tiny ImageNet [26] datasets used in this study are publicly available. The code and result files that support the findings are available from the corresponding author on reasonable request.

REFERENCES

[1] He, K., Zhang, X., Ren, S. and Sun, J., 2016. Deep Residual Learning for Image Recognition. In 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 770-778). https://doi.org/10.1109/cvpr.2016.90

[2] Radenovic, F., Tolias, G. and Chum, O., 2019. Fine-Tuning CNN Image Retrieval with No Human Annotation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 41(7), pp.1655-1668. https://doi.org/10.1109/tpami.2018.2846566

[3] Amaithi Rajan, A., V, V., Raikwar, M. and Balaraman, R., 2024. SMedIR: secure medical image retrieval framework with ConvNeXt-based indexing and searchable encryption in the cloud. Journal of Cloud Computing, 13(1), p.139. https://doi.org/10.1186/s13677-024-00702-z

[4] Tian, M., Zhang, Y., Zhang, Y., Xiao, X. and Wen, W., 2024. A privacy-preserving image retrieval scheme with access control based on searchable encryption in media cloud. Cybersecurity, 7(1), p.22. https://doi.org/10.1186/s42400-024-00213-z

[5] McMahan, B., Moore, E., Ramage, D., Hampson, S. and Aguera y Arcas, B., 2017. Communication-Efficient Learning of Deep Networks from Decentralized Data. In Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, PMLR 54 (pp. 1273-1282).

[6] Phong, L.T., Aono, Y., Hayashi, T., Wang, L. and Moriai, S., 2018. Privacy-Preserving Deep Learning via Additively Homomorphic Encryption. IEEE Transactions on Information Forensics and Security, 13(5), pp.1333-1345. https://doi.org/10.1109/tifs.2017.2787987

[7] Shokri, R., Stronati, M., Song, C. and Shmatikov, V., 2017. Membership Inference Attacks Against Machine Learning Models. In 2017 IEEE Symposium on Security and Privacy (SP) (pp. 3-18). https://doi.org/10.1109/sp.2017.41

[8] Dwork, C. and Roth, A., 2014. The Algorithmic Foundations of Differential Privacy. Foundations and Trends® in Theoretical Computer Science, 9(3-4), pp.211-487. https://doi.org/10.1561/0400000042

[9] Abadi, M., Chu, A., Goodfellow, I., McMahan, H.B., Mironov, I., Talwar, K. and Zhang, L., 2016. Deep Learning with Differential Privacy. In Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security (pp. 308-318). https://doi.org/10.1145/2976749.2978318

[10] Khan, Q., Purification, S. and Chang, S.Y., 2025. Post-Quantum Key Exchange and Subscriber Identity Encryption in 5G Using ML-KEM (Kyber). Information, 16(7), p.617. https://doi.org/10.3390/info16070617

[11] Iavich, M. and Kuchukhidze, T., 2024. Investigating CRYSTALS-Kyber Vulnerabilities: Attack Analysis and Mitigation. Cryptography, 8(2), p.15. https://doi.org/10.3390/cryptography8020015

[12] Bonawitz, K., Ivanov, V., Kreuter, B., Marcedone, A., McMahan, H.B., Patel, S., Ramage, D., Segal, A. and Seth, K., 2017. Practical Secure Aggregation for Privacy-Preserving Machine Learning. In Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security (pp. 1175-1191). https://doi.org/10.1145/3133956.3133982

[13] Evans, D., Kolesnikov, V. and Rosulek, M., 2018. A Pragmatic Introduction to Secure Multi-Party Computation. Foundations and Trends® in Privacy and Security, 2(2-3), pp.70-246. https://doi.org/10.1561/3300000019

[14] Sotthiwat, E., Zhen, L., Li, Z. and Zhang, C., 2021. Partially Encrypted Multi-Party Computation for Federated Learning. In 2021 IEEE/ACM 21st International Symposium on Cluster, Cloud and Internet Computing (CCGrid) (pp. 828-835). https://doi.org/10.1109/ccgrid51090.2021.00101

[15] Zhang, X., Deng, H., Wu, R., Ren, J. and Ren, Y., 2024. PQSF: post-quantum secure privacy-preserving federated learning. Scientific Reports, 14(1), p.23553. https://doi.org/10.1038/s41598-024-74377-6

[16] Hamza, R., 2025. A Quantum-Resistant FHE Framework for Privacy-Preserving Image Processing in the Cloud. Algorithms, 18(8), p.480. https://doi.org/10.3390/a18080480

[17] Sundari, A.N., Sathiya, S. and Gogineni, R.C., 2026. A Comprehensive Survey on Quantum-Resistant Techniques for Content-Based Image Retrieval Systems in Cloud Computing Environments. In Advances in Smart Computing and Emerging Technologies, Lecture Notes in Networks and Systems, vol. 1862 (pp. 369-385). Cham: Springer. https://doi.org/10.1007/978-3-032-19160-1_28

[18] Tang, Z., Fan, H., Ma, H., Chang, S., Gu, X., Fu, Z. and Li, B., 2026. Efficient Privacy-Preserving Image Retrieval With User Revocation in Mobile Cloud. IEEE Transactions on Mobile Computing, 25(4), pp.4791-4805. https://doi.org/10.1109/tmc.2025.3627185

[19] Wang, N., Zhou, W., Wang, J., Guo, Y., Fu, J. and Liu, J., 2024. Secure and Efficient Similarity Retrieval in Cloud Computing Based on Homomorphic Encryption. IEEE Transactions on Information Forensics and Security, 19, pp.2454-2469. https://doi.org/10.1109/tifs.2024.3350909

[20] Chen, L., Yang, Y., Yang, L., Xu, C., Miao, Y., Liu, Z., Liu, X., Choo, K.K.R., Liang, Z. and Hong, C., 2025. Efficient and Secure Content-Based Image Retrieval in Cloud-Assisted Internet of Things. IEEE Internet of Things Journal, 12(5), pp.6001-6013. https://doi.org/10.1109/jiot.2024.3489957

[21] Jia, H., Cai, D., Yang, J., Qian, W., Wang, C., Li, X. and Yang, S., 2023. Efficient and privacy-preserving image classification using homomorphic encryption and chunk-based convolutional neural network. Journal of Cloud Computing, 12(1), p.175. https://doi.org/10.1186/s13677-023-00537-0

[22] Balle, B. and Wang, Y.X., 2018. Improving the Gaussian Mechanism for Differential Privacy: Analytical Calibration and Optimal Denoising. In Proceedings of the 35th International Conference on Machine Learning, PMLR 80 (pp. 394-403).

[23] Belhumeur, P.N., Hespanha, J.P. and Kriegman, D.J., 1997. Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 19(7), pp.711-720. https://doi.org/10.1109/34.598228

[24] Vasilev, V., Shterev, V. and Nenova, M., 2025. Optimizing Activation Function for Parameter Reduction in CNNs on CIFAR-10 and CINIC-10. Applied Sciences, 15(8), p.4292. https://doi.org/10.3390/app15084292

[25] Fei-Fei, L., Fergus, R. and Perona, P., 2004. Learning Generative Visual Models from Few Training Examples: An Incremental Bayesian Approach Tested on 101 Object Categories. In 2004 Conference on Computer Vision and Pattern Recognition Workshop (p. 178). https://doi.org/10.1109/cvpr.2004.383

[26] Hamza, A., Tuama, A. and Mohamed Moubark, A., 2026. FEM-Based Hybrid Compression Framework with Pipeline Implementation for Efficient Deep Neural Networks on Tiny ImageNet. Big Data and Cognitive Computing, 10(5), p.131. https://doi.org/10.3390/bdcc10050131