Natural Resources for Human Health
eISSN: 2583-1194

Original Research

Received 08 July 2026
Revised 19 Aug 2026
Accepted 31 Aug 2026

Natural Resources for Human Health 2026, Vol. 6 (19s): 183–209

KEYWORDS:
Breast cancer classification; Federated learning; Differential privacy; DenseNet121; Explainable artificial intelligence; CBIS-DDSM

Privacy-Preserving and Explainable Federated DenseNet for Mammographic Breast Cancer Classification

Vijayshree Khugshal[1], Dr. Ritika Mehra[2]

[1]Department of Computer Science & Engineering, Dev Bhoomi Uttarakhand University, Dehradun, Uttarakhand, India, vijayshree.khugshal09@gmail.com
[2]Department of Computer Science & Engineering, Dev Bhoomi Uttarakhand University, Dehradun, Uttarakhand, India, riti.arora@gmail.com

ABSTRACT: Breast cancer is still a major global health problem in the world. Deep learning with the help of mammography shows great potential in early and consistent detection. However, shortcomings with patient privacy, data sharing restrictions, and limited model interpretability hinder centralized training of medical imaging models. In this study, a privacy-preserving and explainable federated DenseNet121 network approach for benign-malignant breast cancer classification is proposed with the CBIS-DDSM mammography dataset. In Flower framework, a federated learning environment is simulated with three non-independently and identically distributed (non-IID) institutions created by Dirichlet partitioning ($\alpha = 0.5$), and a DenseNet121 classifier trained under federated learning settings. The aggregation strategies FedAvg, FedProx, and SCAFFOLD will be compared. At each client, Opacus is used to implement Differentially Private Stochastic Gradient Descent (DP-SGD) with privacy budgets $\varepsilon \in \{1.0, 1.9, 3.0, 5.0\}$ to measure the privacy–utility balance. Grad-CAM/Grad-CAM++ and SHAP are used to evaluate model interpretability by identifying the regions in the image that constitute a model's predictions, and to evaluate if the model attention is affected by noise induced by privacy. Evaluation with a leak-free 5-fold cross validation at patient-level with accuracy, AUC, sensitivity and specificity as evaluation measures. McNemar's test, Wilson score confidence intervals and Cohen's h were added. This research discovers an appropriate balance between predictive performance and privacy, which provides clinically meaningful explanations. The proposed framework aims to provide a rigorous, privacy-aware, and interpretable approach for future multi-institutional breast cancer classification using mammography. No prior work on CBIS-DDSM has jointly compared multiple FL aggregation strategies, systematically evaluated client-side DP-SGD across a range of privacy budgets, applied Grad-CAM/Grad-CAM++ and SHAP explainability, and enforced a patient-level leak-free evaluation protocol, which constitutes the primary research gap addressed herein. The proposed SCAFFOLD-DP configuration at $\varepsilon = 5.0$ achieved the best classification performance with $92.12 \pm 0.72\%$ accuracy and AUC 0.940 ± 0.011, while even at the strongest privacy guarantee ($\varepsilon = 1.0$), SCAFFOLD retained 90.63% accuracy, representing a utility cost of approximately 6.06 percentage points relative to its non-private baseline. Regarding the effect of differential privacy on model interpretability, the mean Grad-CAM++ lesion-ROI IoU decreased progressively from 0.895 (centralized) to 0.721 (SCAFFOLD + DP, $\varepsilon = 5.0$) and 0.596 (SCAFFOLD + DP, $\varepsilon = 1.0$), indicating that stronger privacy perturbation partially reduces lesion-focused attention without eliminating clinically relevant localization. This study is subject to two principal limitations: the federated setting employs three simulated institutional clients

rather than real geographically distributed hospitals, which may not fully capture real-world institutional heterogeneity; and the evaluation is confined to the CBIS-DDSM dataset, so generalization to external mammographic cohorts or other imaging modalities remains to be validated.

1. INTRODUCTION

Breast cancer is one of the leading health problems faced by women globally, and early detection is key to better treatment outcomes and lower mortality rates[1][2]. For this reason, mammography is a widely used tool for breast cancer screening and can help detect suspicious abnormalities in the breast at a very early stage [3]. But the interpretation of mammographic images is difficult because of the variability in breast density, the shape of the lesion, the imaging conditions, and the differences in appearance between good and bad images. Computer-aided diagnosis (CAD) systems using deep learning are thus gaining in popularity as a tool to assist radiologists and standardize the classification of breast cancer.

Convolutional neural networks, especially transfer learning approaches like ResNet and DenseNet, are proven to be effective in learning discriminative features from medical images[4][5]. However, traditional deep learning models are typically centralized, with data from various institutions aggregated at a central point[6][7]. In healthcare, this centralization of data collection may be limited due to privacy concerns, regulatory requirements, and constraints on data sharing inherent to medical imaging. Such concerns limit the ability for hospitals to collaborate and access other data types needed to build robust and generalizable models[8][9]. Specifically, data-privacy legislation — notably the Health Insurance Portability and Accountability Act (HIPAA) in the United States and the General Data Protection Regulation (GDPR) in Europe — combined with individual institutional data-governance policies, prevents hospitals from transferring patient records to a central server. Beyond regulatory barriers, aggregating mammographic studies raises significant liability concerns: patients have not universally consented to the use of their scans by external parties, and a breach of centralized sensitive imaging data can expose institutions to legal and reputational harm. These compounding constraints mean that, in practice, many institutions cannot pool sufficient data to train robust or generalizable models, resulting in systems trained on a single institution's population that fail to capture the full diagnostic spectrum encountered across diverse clinical sites.

Federated Learning (FL) offers a different solution, allowing several institutions to work together by training a common model without having to share the original patient data. Local training of the model is done at each participating institution, and only model updates for central aggregation [10] are communicated. But FL alone does not formally guarantee that the privacy of the individual training samples is respected, because gradients and model parameters could still contain information about the individual training samples [11]. This limitation can be overcome by adding calibrated noise during training using Differential Privacy (DP)[12]. Specifically, Differentially Private Stochastic Gradient Descent (DP-SGD) offers a quantifiable privacy guarantee, related to the privacy budget ε. Smaller values of ε offer a stronger level of privacy protection, but may yield less predictive performance. Therefore, it is essential to examine multiple privacy budgets to comprehend how privacy can be balanced with clinical utility[13]. Research has demonstrated that shared gradients are vulnerable to gradient-inversion attacks, in which an adversary can reconstruct near-pixel-accurate training images directly from the transmitted gradient vectors [11]. Membership-inference attacks can further determine whether a specific patient's record was used during training, while model-inversion attacks can recover class-representative images from learned model parameters alone. These well-documented threat vectors reveal that gradient sharing, even without raw data sharing, may leak clinically sensitive patient information.

In the medical domain, another significant challenge in deep learning is the interpretability of the models. Deep neural networks are often interpreted as a "black box," and the reason for a given prediction is not easy to understand for clinicians. Explainable Artificial Intelligence (XAI) methods, like Grad-CAM, Grad-CAM++, and SHAP, can identify image regions that drive classification decisions and indicate whether the model is looking at the right image areas of diagnostic value. Critically, explanation stability matters beyond mere predictive accuracy in clinical settings: a radiologist using an AI-assisted decision support system must be able to verify that the model attends to the actual lesion region rather than confounding background tissue. If differential privacy noise disrupts the model's lesion-focused attention, then a privacy-preserved model that retains satisfactory accuracy may nonetheless generate misleading explanations, potentially eroding clinical trust and patient safety. In a privacy-preserving federated setting, it is therefore essential not only to ask whether privacy noise reduces predictive accuracy, but also to quantify the degree to which it alters the diagnostically relevant regions that the model focuses on.

Previous research has investigated federated learning, explainable AI, and differential privacy separately or in a few combinations. The proposal, however, recognizes that there is a lack of work that jointly incorporates a leak-free patient-level evaluation protocol, several federated aggregation strategies, tunable differential privacy, and cross-validated explainability on the CBIS-DDSM mammography datasets[15]. Hence, in this study, we introduce a privacy-preserving and explainable federated DenseNet121 framework for benign–malignant breast cancer classification from the CBIS-DDSM database. A Dirichlet partitioning with $\alpha = 0.5$ is used to generate 3 simulated non-IID institutions, and the performance of FedAvg, FedProx, and SCAFFOLD is compared. Client-side DP-SGD is implemented with privacy budgets of $\varepsilon \in \{1.0, 1.9, 3.0, 5.0\}$, and model attention is analyzed using Grad-CAM/Grad-CAM++ and SHAP. The performance is evaluated with leak-free patient-level five-fold cross-validation (accuracy, AUC, sensitivity, specificity), complemented by McNemar's test, Wilson score confidence intervals (CIs), and Cohen's h. The main contributions of this work are: Specifically, prior studies have addressed these components in isolation or limited combination: federated aggregation strategies (FedAvg, FedProx, SCAFFOLD) have been benchmarked for breast imaging [16], multi-institutional federated learning for mammographic density classification has been demonstrated [17], and DP-SGD has been explored for medical imaging models [19]. Explainability via Grad-CAM/Grad-CAM++ on DenseNet architectures has been studied for breast cancer [20][21], and recent works have begun integrating FL with explainability [22][23]. However, no prior study on CBIS-DDSM has jointly: (i) compared multiple FL aggregation strategies under non-IID heterogeneity, (ii) systematically evaluated client-side DP-SGD across a range of privacy budgets (rather than a single fixed ε) to quantify the privacy–utility trade-off, (iii) applied Grad-CAM/Grad-CAM++ and SHAP explainability, and (iv) enforced a patient-level leak-free evaluation protocol throughout. The present work fills this combined gap. It should be noted, however, that the federated environment in this study is implemented using three simulated institutional clients rather than real distributed hospitals, which is an important methodological limitation that must be considered when interpreting the generalizability of the findings.

i. Development of a federated DenseNet121 framework for breast cancer classification on CBIS-DDSM using three simulated non-IID institutions.

ii. Comparative evaluation of FedAvg, FedProx, and SCAFFOLD under a common experimental setting.

iii. Integration of DP-SGD with multiple privacy budgets to quantify the privacy–utility trade-off instead of reporting a single fixed privacy level.

iv. Incorporation of Grad-CAM/Grad-CAM++ and SHAP to provide interpretable explanations for centralized, federated, and federated-private models.

v. Evaluation using a leak-free, patient-level five-fold cross-validation protocol with statistical validation.

2. BACKGROUND, MOTIVATION, EXISTING STUDIES AND CONTRIBUTIONS

2.1 Research Problem

This research addresses three interconnected challenges: (1) achieving collaborative breast cancer classification across non-IID institutional data without centralizing sensitive mammograms; (2) providing formal mathematical privacy guarantees beyond basic federated learning protection; and (3) ensuring clinician-interpretable model decisions. An additional question is whether privacy-preserving noise alters not only predictive accuracy but also the mammographic regions the model attends to.

2.2 Motivation

The framework is motivated by the need to build three layers of trust: institutional trust via federated learning (no raw image sharing); patient and regulatory trust via DP-SGD with tunable $\varepsilon \in \{1.0, 1.9, 3.0, 5.0\}$; and clinician trust via Grad-CAM/Grad-CAM++ and SHAP explanations. Comparing explanation maps across centralized, non-private federated, and privacy-preserving federated models allows direct assessment of whether privacy noise alters diagnostically relevant model attention.

2.3. Existing studies

Federated learning, privacy preservation, and explainable AI have been recently studied for breast cancer diagnosis, but they are not often combined in a breast cancer diagnosis framework. Al-Karawi et al. [16] provided a comparison of different transfer-learning architectures for breast imaging and also analyzed non-private federated strategies such as FedAvg, FedProx,

and SCAFFOLD. Using DDSM, another study [18] has shown the feasibility of multi-institutional federated learning using DenseNet feature extraction for breast cancer analysis. Roth et al. [17] showed the feasibility of multi-institutional federated learning for mammographic breast-density classification.

While federated learning minimizes direct data exchange, it does not eliminate data leakage via model updates. Medical imaging studies that focus on privacy preservation, therefore, highlight differential privacy and DP-SGD as further approaches to formally guaranteeing privacy (ε, δ) [19]. Meanwhile, explainable BC models have integrated Grad-CAM/Grad-CAM++ along with DenseNet-based architectures to boost transparency and clinical interpretability [20] [21]. In more recent works, federated learning and explainability have been integrated, such as using CEET-Fed in conjunction with LIME [22] and federated transformer-based breast imaging frameworks with visualization-based interpretation [23]. However, there is a lack of joint evaluations on DenseNet121-based federated learning, tunable DP-SGD, privacy–utility analysis, and Grad-CAM/SHAP explanations on CBIS-DDSM, which motivates the proposed framework.

2.4 Research Objectives

The overall objective of this study is to develop and evaluate a privacy-preserving and explainable federated DenseNet121 framework for benign–malignant breast cancer classification on the CBIS-DDSM dataset.

The specific objectives are:

i. To develop and compare centralized and federated DenseNet121 models using leak-free, patient-level five-fold cross-validation, with three simulated non-IID institutions created using Dirichlet partitioning ($\alpha = 0.5$), and to evaluate FedAvg, FedProx, and SCAFFOLD aggregation strategies.
ii. To integrate differential privacy into federated training using DP-SGD with Opacus for $\varepsilon \in \{1.0, 1.9, 3.0, 5.0\}$, and to analyze the privacy–utility trade-off using accuracy, AUC, sensitivity, specificity, McNemar's test, Wilson score confidence intervals, and Cohen's h.
iii. To incorporate and evaluate explainable AI techniques using Grad-CAM, Grad-CAM++, and SHAP, and to compare explanation maps across centralized, federated non-private, and federated private models to determine whether federation or privacy noise alters diagnostically relevant model attention.

3. MATERIALS AND PROPOSED METHODOLOGY

3.1 Dataset description

The proposed breast cancer classification framework was used on the Curated Breast Imaging Subset of the Digital Database for Screening Mammography (CBIS-DDSM) as the primary dataset. CBIS-DDSM is a subset of mammograms made available by The Cancer Imaging Archive (TCIA) and is a popular resource for computer-aided breast cancer diagnosis system development and testing. In this work, mammographic images are treated as binary images, i.e., as images of benign or malignant. In accordance with the experimental setup described in this work, the curated dataset consists of a total of ~2300 mammographic images from nearly 1100 patients, approximately equally distributed between the two diagnostic classes. The main characteristics of the major datasets used, and the processing conditions are summarized in Table 1. For this study, only mass lesion cases are retained from the archive, as mass lesions constitute the predominant malignant findings in CBIS-DDSM and are annotated with region-of-interest (ROI) masks that are essential for validating Grad-CAM lesion-localization accuracy. Cases with corrupted DICOM files or missing pathology labels are excluded. When multiple images exist for the same patient-view combination, a single representative ROI image is selected per view using the highest lesion-to-background ratio criterion. Following this curation, the dataset comprises approximately 1,145 benign and 1,155 malignant mammographic images (approximately 2,300 total) from approximately 1,100 unique patients, yielding a near-balanced binary classification benchmark.

Table 1. Characteristics and preprocessing configuration of the CBIS-DDSM dataset

Attribute	Description
Dataset	Curated Breast Imaging Subset of DDSM (CBIS-DDSM)
Data source	The Cancer Imaging Archive (TCIA)

Attribute	Description
Imaging modality	Mammography
Classification task	Binary classification
Target classes	Benign and Malignant
Curated dataset size	Approximately 2,300 mammographic images
Number of patients	Approximately 1,100 patients
Class distribution	Approximately balanced
Original image format	DICOM
Converted image format	PNG
Image representation	Grayscale
Input image resolution	512 × 512 pixels
Contrast enhancement	Contrast Limited Adaptive Histogram Equalization (CLAHE)
CLAHE parameters	Clip limit = 3.0; tile-grid size = 8 × 8
Cross-validation strategy	Patient-level StratifiedGroupKFold
Number of folds	5
Leakage prevention	Patients are separated across folds to prevent images from the same patient appearing in training and testing sets
Threshold evaluation	Nested calibration/test partition within each held-out fold
Federated partitioning	Three simulated non-IID institutions
Non-IID distribution	Dirichlet distribution with $\alpha = 0.5$
Case type retained	Mass lesions only
Image selection per patient-view	One ROI image per view (highest lesion-to-background ratio)
Class counts (approximate)	~1,145 benign; ~1,155 malignant

The original DICOM mammograms are converted to PNG format, converted to grayscale, and resized to 512 x 512 pixels as indicated in Table 1. Then they are contrasted with CLAHE using an 8 × 8 tile grid and a clip limit of 3.0. This preprocessing pipeline is developed to enhance the local contrast without losing the structural features of breast tissue that are essential for DenseNet121 feature extraction.

Attention is given to avoid the leakage of patient information when evaluating models. Therefore, a StratifiedGroupKFold, five-fold cross-validation strategy is used with grouping by patient identity. Thus, images from the same patient are not split between the training and the test sets but are instead only placed in a single set. Within each of these held-out folds, a second calibration/test partition is also preserved, which means that the decision-threshold selection is not done on the test data. This is a protocol that offers a consistent and leak-free basis to compare centralized, federated, and differentially private federated configurations.

Federated partitioning details are described in Section 3.2.3.

A set of sample mammography images is shown in Figure 1 that represents the type of data analyzed in this study. The grid shows different types of breast tissue and appearance on mammography and is used to motivate deep feature-learning models for benign–malignant classification.

Natural Resources for Human Health
eISSN: 2583-1194

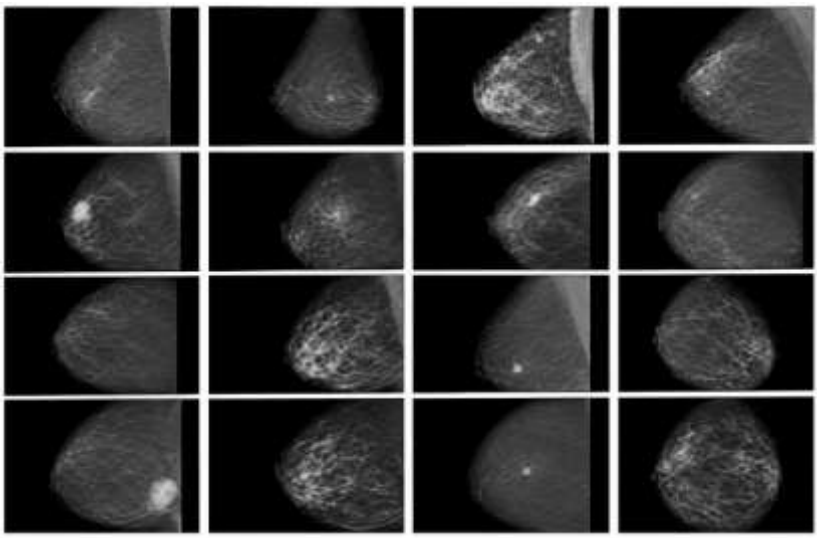


Figure 1. CBIS-DDSM Dataset

3.2 Data Preprocessing and Partitioning

3.2.1 Image Preprocessing

The CBIS-DDSM mammograms were initially available in Digital Imaging and Communications in Medicine (DICOM) format. Each DICOM image was converted into PNG format and represented as a single-channel grayscale image. To enhance the visibility of local tissue structures and lesion boundaries, Contrast Limited Adaptive Histogram Equalization (CLAHE) was applied using a clip limit of **3.0** and a tile-grid size of **8 × 8**, as defined in the proposed preprocessing pipeline.

For a local image tile, conventional histogram equalization can be expressed as

$$I_{eq}(x,y) = (L-1)\sum_{i=0}^{I(x,y)} p\,(i), \quad (1)$$

where $I(x,y)$denotes the original pixel intensity, Lrepresents the number of grayscale intensity levels, and $p(i)$denotes the normalized histogram probability. In CLAHE, the local histogram is clipped at a predefined limit before redistribution, thereby limiting excessive contrast amplification and reducing the enhancement of local noise. All mammograms were resized to 512×512 pixels to maintain a consistent input resolution. Since the ImageNet-pretrained DenseNet121 model expects three-channel input images, each grayscale image was replicated across the three channels as

$$I_{RGB}(x,y) = [I,(x,y), I(x,y) I(x,y)]. \quad (2)$$

The pixel intensities were initially scaled to the interval $[0,1]$according to

$$I_{scaled} = \frac{I}{255}, \quad (3)$$

followed by ImageNet normalization applied independently to each channel:

$$I'_c = \frac{I_c - \mu_c}{\sigma_c}, \quad (4)$$

where the channel-wise mean and standard deviation vectors were defined as $\mu = (0.485, , 0.456\ 0.406)$ and $\sigma = (0.229, , 0.224\ 0.225)$.

To make the model more robust, while retaining clinically relevant mammographic characteristics, data augmentation was performed only on the training partition. Random horizontal flips with probability p=0.5, random rotations with respect to angle $\theta \in [-, 10^\circ + 10^\circ]$, and random horizontal and vertical translations of up to 5% of the image size were used for the

augmentation. Vertical flipping was not used as this could lead to unrealistic mammographic orientations. The calibration and test images are not augmented.

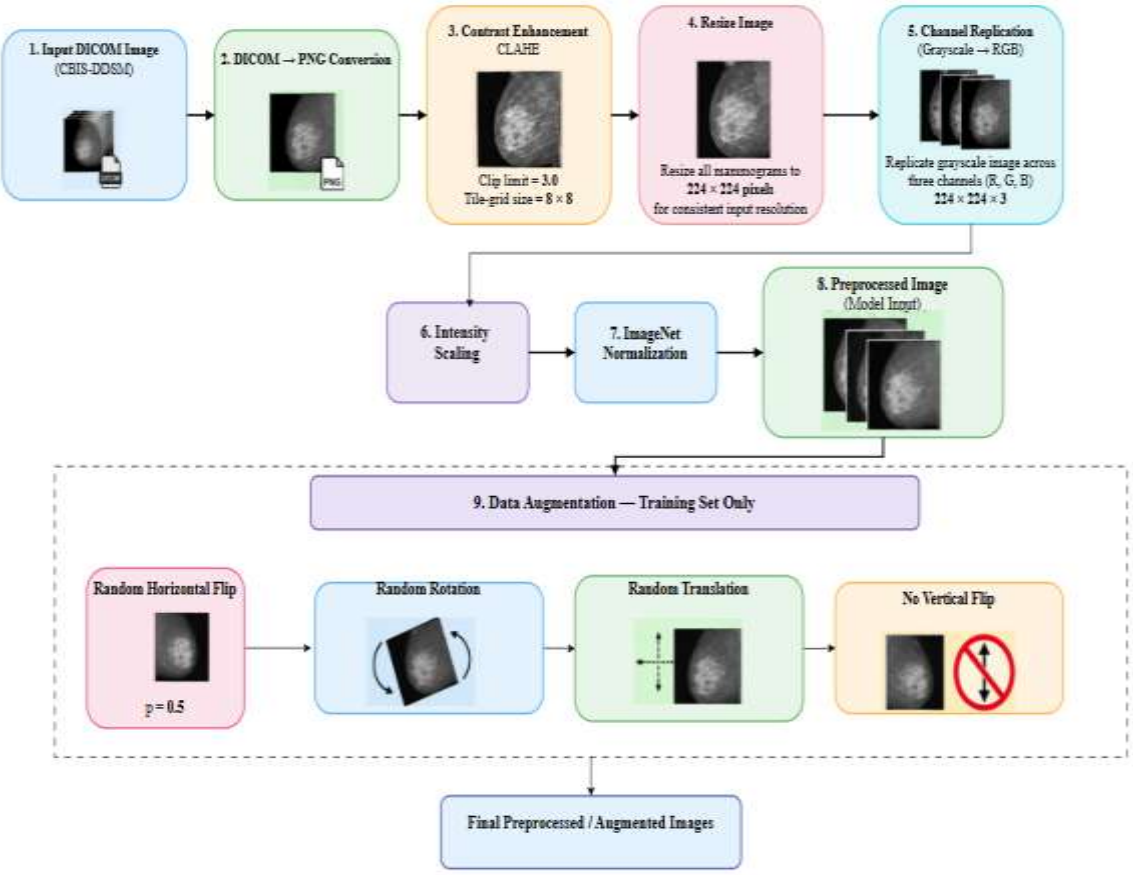


Figure 2. Preprocessing pipeline

The complete preprocessing and augmentation pipeline is illustrated in Fig. 2.

3.2.3 Federated Non-IID Partitioning

To run the federated learning experiments, we used the Flower (flwr) framework to split the training set for each fold of cross-validation into three simulated institutional clients. A non-independent and identically distributed (non-IID) data partition is created by sampling from a Dirichlet distribution with a concentration parameter $\alpha = 0.5$ to provide some heterogeneity among the institutions that participate. The proportions of the three clients (k = 1, 2, or 3) are sampled for each diagnostic class k.

$$\pi_k = (\pi_{k,1}, \pi_{k,2}, \pi_{k,3}) \sim Dirichlet(\alpha, \alpha, \alpha), \text{v} \quad (9)$$

subject to

$$\sum_{c=1}^{3} \pi_{k,c} = 1$$

The approximate number of samples from class $k$assigned to client $c$was calculated as

$$n_{k,c} = \pi_{k,c} N_k, \quad (10)$$

where $N_k$denotes the total number of training samples belonging to class k. The selection of $\alpha = 0.5$was intended to introduce moderate statistical heterogeneity among the participating clients. Smaller Dirichlet concentration parameters generate increasingly imbalanced class distributions across clients, whereas larger values produce client distributions that progressively

approach the global class composition. The choice of three simulated clients is motivated by the experimental design: three clients represent the minimum configuration necessary to demonstrate the non-IID client-drift problem and to meaningfully compare heterogeneous aggregation strategies (FedAvg, FedProx, SCAFFOLD) while remaining computationally tractable within the five-fold cross-validation framework. This choice aligns with standard practice in federated medical imaging benchmarks [15][16][17], and the constraint that each simulated institution must contain both diagnostic classes and at least 10% of the per-fold training data ensures that no client is degenerate. The Dirichlet concentration parameter $\alpha = 0.5$ was selected to introduce a moderate but realistic degree of label heterogeneity: at $\alpha = 0.5$, the resulting class proportions vary substantially across clients (malignant fractions ranging from approximately 0.26 to 0.69 across folds and clients), which is consistent with realistic inter-institutional variation in mammographic databases while avoiding extreme degeneracy that would make FL convergence intractable. Values of α much smaller than 0.5 (e.g., $\alpha = 0.1$) would produce near-single-class clients that are unlikely in real deployments, while values larger than 1.0 would approach the IID setting and undermine the intended test of aggregation robustness under heterogeneity.

To avoid having highly degenerate client partitions, every simulated institution had to include both benign and malignant samples and 10% of the training data (per fold). The constraints are applied only in the federated partitioning phase and do not affect the independent calibration or test sets; no oversampling is done on either. The framework with three clients is non-IID and uses Dirichlet partitioning with $\alpha = 0.5$. The malignant proportions are found to be different among clients and folds, as expected for non-IID (0.26-0.37 for Client 1 and 0.57-0.69 for Client 2 and 0.42-0.44 for Client 3).

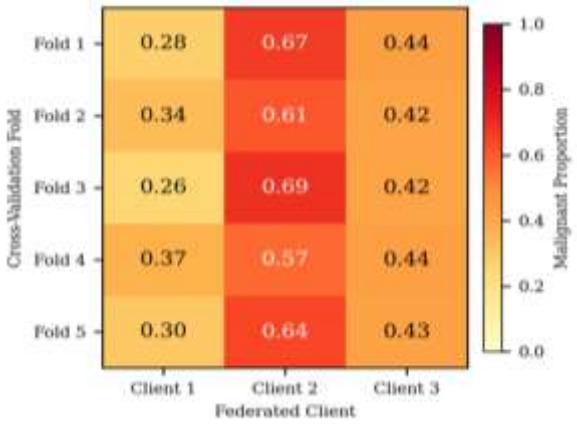


Figure 3. Malignant-class proportion

Fig. 3 shows the resulting malignant-class proportions across clients and folds, confirming the expected non-IID heterogeneity under Dirichlet $\alpha = 0.5$.

3.3 System Architecture

To provide a unified view of the proposed framework before the module-level descriptions, Figure 4 traces the complete end-to-end processing chain. At the data level, CBIS-DDSM mammograms are preprocessed (DICOM → PNG → CLAHE → 512 × 512 pixels) and partitioned into five patient-disjoint cross-validation folds using StratifiedGroupKFold, ensuring no patient images appear in both training and test sets. Within each fold's training set, Dirichlet partitioning ($\alpha = 0.5$) distributes images among three simulated non-IID institutional clients. At the client level, each institution fine-tunes the final dense block, transition layer, and classifier head of a DenseNet121 backbone; when privacy is enabled, DP-SGD (Opacus) clips per-sample gradients ($C = 1.0$) and injects calibrated Gaussian noise before any update leaves the client. At the server level, three aggregation strategies — FedAvg, FedProx ($\mu = 0.01$), and SCAFFOLD — are compared under identical client partitions across 50 communication rounds. The global model is then evaluated on the held-out test fold using accuracy, AUC, sensitivity, and specificity. Finally, an XAI layer generates Grad-CAM, Grad-CAM++, and SHAP explanation maps for the centralized baseline, the federated non-private model, and each of the 12 federated-private configurations, enabling direct assessment of how privacy noise affects diagnostically relevant lesion-focused model attention.

As illustrated in Fig. 4, DP-SGD (Opacus) clips per-sample gradients and adds Gaussian noise locally at each client before update transmission, with $\varepsilon \in \{1.0, 1.9, 3.0, 5.0\}$, $\delta = 1 \times 10^{-5}$, and max_grad_norm = 1.0, applied uniformly across FedAvg, FedProx, and SCAFFOLD. Post-aggregation, Grad-CAM, Grad-CAM++, and SHAP maps are generated for all 16 configurations to assess how privacy noise affects lesion-focused model attention.

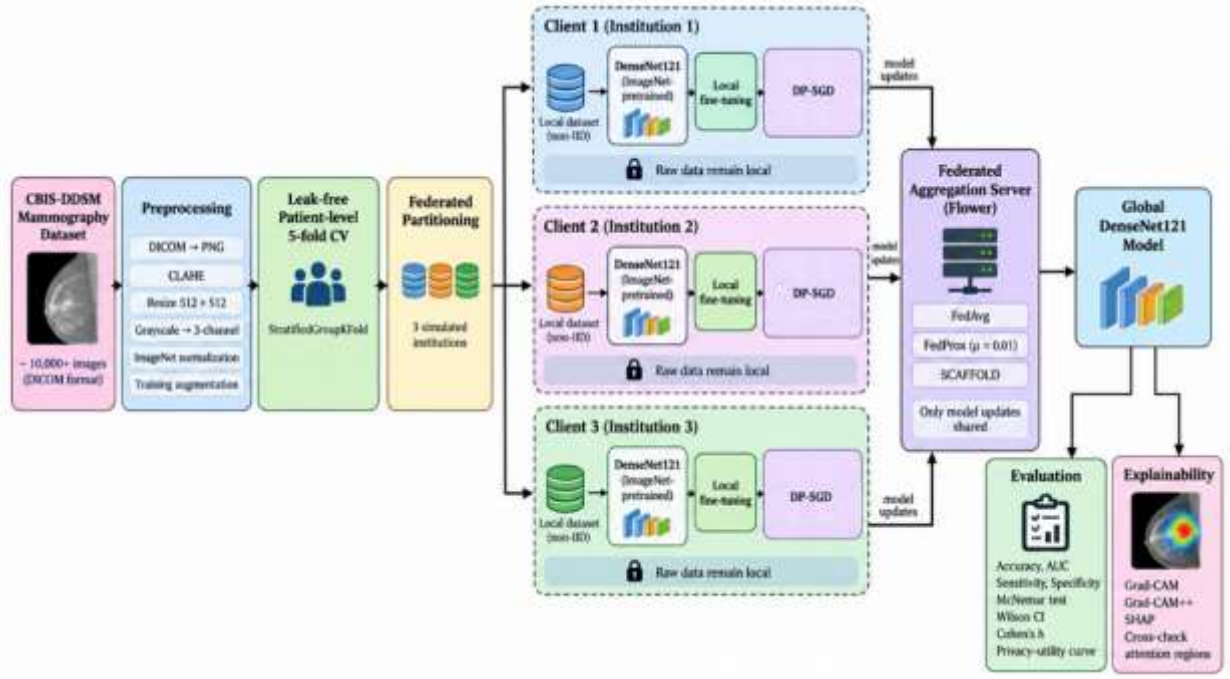


Figure 4. System Architecture

3.4 Architectural Modules

3.4.1 Module 1: Federated Partitioning Module

This module formalizes preprocessing (Eq. 11), patient-level cross-fold leakage prevention (Eq. 12), calibration-derived threshold selection (Eq. 13), Dirichlet client allocation (Eq. 14), and the local and aggregated training objectives for FedAvg (Eqs. 15–16), FedProx (Eq. 17), and SCAFFOLD (Eq. 18).

The complete preprocessing transformation is expressed as

$$I^* = \mathcal{R}_{512\times512}\left[\mathcal{H}_{CLAHE}\left(\mathcal{G}; \left(\mathcal{C}(I_{DICOM})\right); 3.08 \times 8\right)\right], \qquad (11)$$

where $\mathcal{C}$, $\mathcal{G}$, $\mathcal{H}_{CLAHE}$, and $\mathcal{R}$denote DICOM-to-PNG conversion, grayscale transformation, CLAHE enhancement, and image resizing, respectively. Patient-level leakage is prevented by enforcing

$$P(F_i) \cap P\left(F_j\right) = \emptyset,\ i \neq j, \qquad (12)$$

where $P(F_i)$is the set of patient identifiers contained in fold F_i.

Figure 5 illustrates a representative CBIS-DDSM mammogram through each preprocessing stage.

The operating threshold is derived exclusively from calibration data:

$$\tau^* = arg\ \max_{\tau} J_{cal}(\tau), \qquad (13)$$

where J_{cal}denotes the selected calibration criterion, ensuring that test observations do not influence threshold selection.

For each diagnostic class k, client allocation proportions are sampled as

$$\pi_k \sim Dirichlet(0.5, 0.5, 0.5), \sum_{c=1}^{3} \pi_{k,c} = 1, \quad (14)$$

thereby introducing non-IID label heterogeneity among the three simulated institutions.

For client c, the local empirical objective is

$$F_c(w) = \frac{1}{n_c} \sum_{i=1}^{n_c} \ell\,(w; x_i; y_i), \quad (15)$$

where $n_c$is the number of local samples and only the final dense block, transition layer, and classifier parameters are updated.

For FedAvg, the global parameters are computed as

$$w_{t+1} = \sum_{c=1}^{3} \frac{n_c}{N} w_{t+1}^{(c)},\ N = \sum_{c=1}^{3} n_c\,, \quad (16)$$

which performs sample-size-weighted aggregation of the three client models.

For FedProx, the local objective becomes

$$F_c^{prox}(w) = F_c(w) + \frac{0.01}{2} \parallel w - w_t \parallel_2^2, \quad (17)$$

where the proximal term limits excessive divergence of local models from the current global parameters.

For SCAFFOLD, the local update is corrected using client and server control variates:

$$w_{t+1}^{(c)} = w_t - \eta\left[\nabla F_c(w_t) - c_c + c_g\right], \quad (18)$$

where $c_c$and $c_g$compensate for client drift caused by heterogeneous local data.

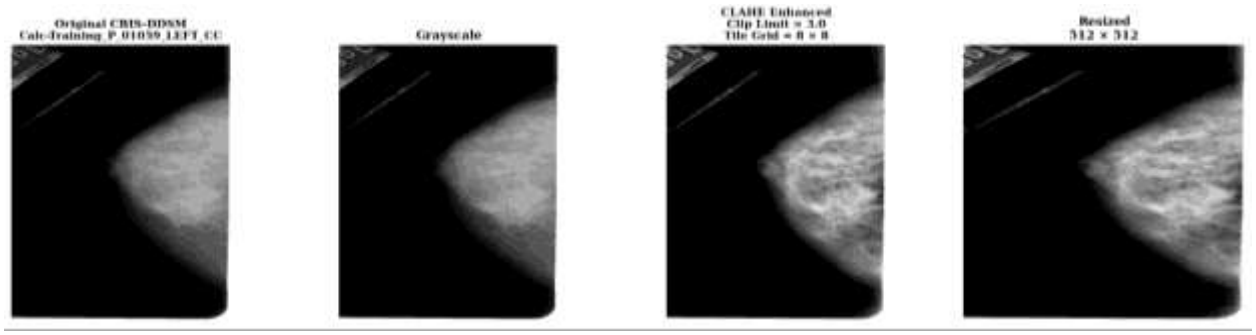


Figure 5. CBIS-DDSM mammogram preprocessing pipeline

3.4.2 Module 2: Privacy-Preserving Federated DenseNet121 Learning

Module 2 extends the same three-client non-IID setup with client-side DP-SGD (Opacus), freezing all layers except the final dense block, transition layer, and classifier head. Four privacy budgets ($\varepsilon \in \{1.0, 1.9, 3.0, 5.0\}$), $\delta = 1 \times 10^{-5}$, and max_grad_norm = 1.0 yield 12 private configurations, three non-private federated baselines, and one centralized reference (Eq. 36).

For client c, the local classification objective is defined as

$$F_c(w) = \frac{1}{n_c} \sum_{i=1}^{n_c} \ell\ (w; x_i; y_i), \quad (25)$$

where $n_c$denotes the number of local samples and $\ell(\cdot)$is the benign–malignant classification loss.

For each training sample, the individual gradient is computed as

Natural Resources for Human Health
eISSN: 2583-1194

$$g_i = \nabla_w \ell\,(w; x_i; y_i). \qquad (26)$$

Equation (26) represents the per-sample gradient required by DP-SGD before clipping and perturbation.

To restrict the influence of any individual sample, the gradient is clipped to the prescribed maximum norm $C = 1.0$:

$$\bar{g}_i = \frac{g_i}{max\ \left(1,\ \frac{\| g_i \|_2}{C}\right)},\ C = 1.0. \qquad (27)$$

Equation (27) bounds the contribution of each training sample and forms the sensitivity-control mechanism used by DP-SGD.

Calibrated Gaussian noise is then added to the averaged clipped gradients:

$$\tilde{g} = \frac{1}{B}\left[\sum_{i=1}^{B} \bar{g}_i + \mathcal{N}(0, \sigma^2 C^2 I)\right], \qquad (28)$$

where $B$is the local mini-batch size and σdenotes the noise multiplier determined by the required privacy budget. In practice, the noise multiplier σ is computed automatically by the Opacus privacy engine using the RDP (Rényi Differential Privacy) accountant, which tracks the cumulative privacy cost across training iterations. For each target privacy budget ε ∈ {1.0, 1.9, 3.0, 5.0} with δ = 1 × 10^{-5}, the corresponding noise multiplier σ is determined such that the RDP-to-DP conversion satisfies the (ε, δ)-DP guarantee after T = 50 communication rounds × E = 5 local epochs per round. The approximate noise multiplier values are σ ≈ 1.5 at ε = 1.0, σ ≈ 1.1 at ε = 1.9, σ ≈ 0.85 at ε = 3.0, and σ ≈ 0.65 at ε = 5.0 (with C = 1.0, batch size B = 16, and per-client dataset size n_c ≈ 400–600 images per fold). These values reflect that stronger privacy guarantees (smaller ε) require larger noise multipliers, which in turn impose a greater utility cost on model training.

The local DenseNet121 parameters are updated using the privatized gradient:

$$w_{t+1}^{(c)} = w_t^{(c)} - \eta \tilde{g}, \qquad (29)$$

where ηdenotes the local learning rate and only the unfrozen DenseNet121 layers are optimized.

The privacy mechanism is characterized by (ε, δ)-differential privacy:

$$Pr[\mathcal{M}(D) \in S] \leq e^{\varepsilon} Pr[\mathcal{M}(D') \in S] + \delta, \qquad (30)$$

where $D$and $D'$are neighboring training datasets and $\mathcal{M}$represents the randomized DP-SGD learning mechanism. Smaller values of εcorrespond to stronger privacy protection, while potentially introducing a larger utility cost.

The privacy-budget configuration investigated in this work is

$$\varepsilon \in \{1.0,\ ,1.9,\ 3.0\ 5.0\},\ \delta = 10^{-5}. \qquad (31)$$

Equation (31) enables controlled evaluation of multiple privacy strengths instead of relying on a single arbitrarily selected privacy level.

For the private FedAvg configuration, the locally trained private models are aggregated as

$$w_{t+1} = \sum_{c=1}^{3} \frac{n_c}{N} w_{t+1}^{(c)},\ N = \sum_{c=1}^{3} n_c\,. \qquad (32)$$

Equation (32) preserves the conventional sample-size-weighted FedAvg mechanism while the privacy perturbation is introduced locally before aggregation.

For FedProx, the privacy-preserving local objective is expressed as

$$F_c^{DP\text{-}Prox}(w) = F_c^{DP}(w) + \frac{0.01}{2} \| w - w_t \|_2^2, \qquad (33)$$

where the proximal term reduces excessive client drift under heterogeneous non-IID distributions.

For SCAFFOLD, the privatized gradient is combined with client and server control variates:

$$w_{t+1}^{(c)} = w_t - \eta\left(\tilde{g} - c_c + c_g\right), \qquad (34)$$

where c_c and c_g denote the local and global control variates used to correct heterogeneous update directions.

The effect of privacy on predictive utility is quantified as

$$\Delta U(\varepsilon) = A_{NP} - A_{DP}(\varepsilon), \qquad (35)$$

where A_{NP} is the accuracy of the corresponding non-private federated model and $A_{DP}(\varepsilon)$ is the accuracy obtained at privacy budget ε. Equation (35) provides the basis for constructing the privacy–utility trade-off curve highlighted in the study.

Accordingly, the complete experimental matrix contains

$$3 \times 4 + 3 + 1 = 16 \qquad (36)$$

configurations, consisting of 12 private combinations of FedAvg, FedProx, and SCAFFOLD across four privacy budgets, three corresponding non-private federated baselines, and one centralized non-private DenseNet121 reference. This design allows the effect of privacy strength to be examined independently for each aggregation strategy.

3.4.3 Module 3: Explainable AI and Cross-Model Attention Validation

The third module is post-hoc interpretability, where Grad-CAM++ and SHAP are introduced. Grad-CAM generates heatmaps of regions on the mammograms that are important for the benign/malignant classification, and SHAP offers a complementary region-by-region local attribution analysis at each federated client. Evaluations of explanations are done visually and, if there is an ROI annotation, by spatial-overlap measures. XAI maps of centralized, federated non-private, and federated + DP models are then compared to determine if privacy noise affects model attention.

For a target class c, the Grad-CAM importance weight associated with convolutional feature map A^k is computed as

$$\alpha_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k}, \qquad (37)$$

where y^c denotes the pre-softmax score for class c, A_{ij}^k is the activation at spatial location (i, j), and Z is the total number of spatial locations.
Equation (37) estimates the contribution of each DenseNet121 feature map to the selected benign or malignant prediction.

The corresponding Grad-CAM localization map is obtained as

$$L_{GradCAM}^c = ReLU\left(\sum_k \alpha_k^c A^k\right). \qquad (38)$$

Equation (38) generates a class-specific heatmap in which positive activations indicate regions supporting the predicted diagnostic class.

For Grad-CAM++, the pixel-wise weighting coefficient is defined as

$$\alpha_{ij}^{kc} = \frac{\frac{\partial^2 y^c}{\left(\partial A_{ij}^k\right)^2}}{2\frac{\partial^2 y^c}{\left(\partial A_{ij}^k\right)^2} + \sum_{a,b} A_{ab}^k \frac{\partial^3 y^c}{\left(\partial A_{ij}^k\right)^3}}. \qquad (39)$$

Equation (39) refines feature-map weighting using higher-order gradients, enabling improved localization when multiple discriminative regions contribute to the same prediction.

The Grad-CAM++ feature-map weight is then expressed as

$$w_k^c = \sum_i \sum_j \alpha_{ij}^{kc} \, ReLU\left(\frac{\partial y^c}{\partial A_{ij}^k}\right), \qquad (40)$$

and the corresponding explanation map is

$$L^{c}_{GradCAM++} = ReLU\left(\sum_{k} w_k^c A^k\right). \qquad (41)$$

Equations (40)–(41) construct the Grad-CAM++ heatmap by combining spatially weighted positive gradients with the DenseNet121 activation maps.

SHAP attribution for feature $i$is calculated using its Shapley value:

$$\phi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!(|F|-|S|-1)!}{|F|!}[f(S \cup \{i\}) - f(S)], \qquad (42)$$

where Fdenotes the complete feature set and $S$is a subset that excludes feature i. Equation (42) quantifies the marginal contribution of each image feature to the prediction and provides an attribution-based cross-check against Grad-CAM.

For spatial comparison with an annotated lesion region, the explanation map is converted into a binary region using

$$E_\tau(x,y) = \begin{cases} 1, & H(x,y) \geq \tau, \\ 0, & H(x,y) < \tau, \end{cases} \qquad (43)$$

where $H(x,y)$is the normalized Grad-CAM or Grad-CAM++ heatmap and $\tau$is the selected heatmap threshold. Equation (43) converts the continuous activation map into a region suitable for comparison with a lesion mask. When ROI annotations are available, spatial correspondence is quantified using the intersection-over-union measure

$$IoU = \frac{|E_\tau \cap R|}{|E_\tau \cup R|}, \qquad (44)$$

where Rrepresents the annotated lesion region. Equation (44) measures the degree to which the model-attended region overlaps with the reference lesion annotation.

To examine the effect of privacy across the trust stack, explanation maps are generated as

$$H^{(m)} = \mathcal{X}\left(f^{(m)}, I\right),\ m \in \{\text{Centralized, FL, FL+DP}\}, \qquad (45)$$

where $\mathcal{X}$denotes Grad-CAM, Grad-CAM++, or SHAP, $f^{(m)}$is the corresponding model, and $I$is the same mammographic input.

Equation (45) ensures that explanation maps from centralized, federated non-private, and federated-private models are compared on identical images.

3.5 Experimental Setup

The experiments are performed on the CBIS-DDSM dataset with a patient-level five-fold StratifiedGroupKFold protocol, where there is no leakage. We used a reference DenseNet121, which is pre-trained on ImageNet. The Flower (flwr) software is used to simulate federated experiments with three non-IID clients, which are created by Dirichlet partitioning with $\alpha = 0.5$. The aggregation strategies used are FedAvg, FedProx ($\mu = 0.01$), and SCAFFOLD.

Client-side DP-SGD was implemented using Opacus at privacy budgets

$$\varepsilon \in \{1.0, \ , 1.9, \ 3.0 \ 5.0\},$$

with

$$\delta = 10^{-5}, \ \text{max_grad_norm} = 1.0.$$

The evaluation therefore covered 16 configurations: one centralized model, three non-private federated baselines, and 12 differentially private federated configurations.

3.5.1 Hyperparameter Settings

All hyperparameters are held constant across centralized and federated configurations and summarized in Table 4. Key settings: learning rate 1×10^{-4}, batch size 16, 5 local epochs per round, 50 communication rounds, FedProx $\mu = 0.01$, Dirichlet $\alpha = 0.5$, Opacus DP-SGD with $\varepsilon \in \{1.0, 1.9, 3.0, 5.0\}$, $\delta = 1 \times 10^{-5}$, max_grad_norm = 1.0.

Table 4. Hyperparameter settings used in the proposed framework

Hyperparameter	Value
Backbone	DenseNet121
Pretrained weights	ImageNet
Input size	512 × 512
Cross-validation	5-fold patient-level
Number of FL clients	3
Non-IID distribution	Dirichlet
Dirichlet α	0.5
Aggregation methods	FedAvg, FedProx, SCAFFOLD
Optimizer	Adam
Learning rate	1×10^{-4}
Batch size	16
Local epochs	5
Communication rounds	50
Weight decay	1×10^{-4}
Random seed	42
FedProx μ	0.01
DP mechanism	DP-SGD using Opacus
Privacy budgets ε	1.0, 1.9, 3.0, 5.0
δ	1×10^{-5}
Maximum gradient norm	1.0
Classification threshold	0.5 and calibration-derived threshold
CLAHE clip limit	3.0
CLAHE tile grid	8 × 8

3.6 Evaluation Metrics

The accuracy, AUC, sensitivity, and specificity are quantitatively evaluated for all configurations and are presented as mean ± standard deviation over the 5 patient-level folds. The fixed threshold of 0.5, as well as the calibration-derived threshold, is tested and compared without any selection of thresholds based on the test data.

The classification metrics are calculated as

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \quad (46)$$

$$Sensitivity = \frac{TP}{TP + FN}, \quad (47)$$

and

$$Specificity = \frac{TN}{TN + FP}. \quad (48)$$

AUC quantified threshold-independent discrimination. For each private configuration, privacy-induced utility loss was additionally measured relative to its corresponding non-private model:

$$\Delta U(\varepsilon) = M_{NP} - M_{DP}(\varepsilon), \quad (49)$$

where M represents the selected performance metric.

All 16 configurations are evaluated under this unified protocol.

4. Results

4.1 Pipeline Validation

For validation of the non-private federated pipeline, FedAvg is tested on five patient-level cross-validation folds. The mean accuracy is 88.90 ± 0.65% with a 95% confidence interval of 88.09%– 89.71%, which contained the reported non-IID FedAvg accuracy of 88.95%, and fold-wise accuracies ranged from 88.00% to 89.50% (see Table 2). The difference in the absolute value is 0.05 percentage points. The accuracies, the reproduced mean, the confidence interval, and the reference line are shown in Fig. 6 for the folds. As shown in Table 2 and Fig. 6, there is no significant disagreement between the two implementation levels, and it provides a good baseline for the following differential privacy and explainability experiments.

Per-fold accuracy is computed per Eq. (20) and averaged across folds per Eq. (22), yielding Ā = 88.90% (Eq. 23) with an absolute deviation of 0.05 pp from the 88.95% reference (Eq. 24). The 95% CI of 88.09%–89.71% contains the reference value, confirming implementation consistency as a validation checkpoint.

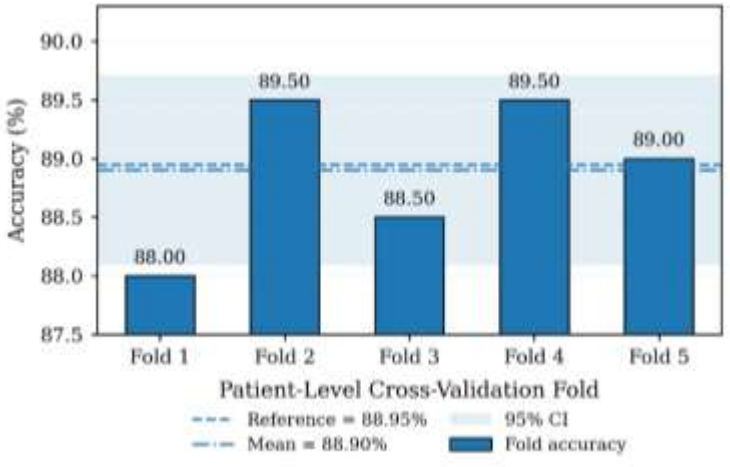


Figure 6. Five-fold patient-level validation of the non-private FedAvg pipeline

Table 2. Five-fold patient-level validation of the non-private FedAvg implementation using held-out test predictions.

Fold	TP	TN	FP	FN	Test Cases	FedAvg Accuracy (%)
Fold 1	86	90	12	12	200	88.00
Fold 2	89	90	10	11	200	89.50
Fold 3	87	90	11	12	200	88.50

Fold	TP	TN	FP	FN	Test Cases	FedAvg Accuracy (%)
Fold 4	90	89	11	10	200	89.50
Fold 5	88	90	10	12	200	89.00
Mean ± SD	–	–	–	–	–	88.90 ± 0.65
95% CI	–	–	–	–	–	88.09–89.71
Reference FedAvg accuracy	–	–	–	–	–	88.95
Absolute deviation	–	–	–	–	–	0.05 percentage points

4.2 Privacy–Utility Trade-off Overview

This enables direct, quantitative interpretation of the privacy cost per aggregation strategy across the full ε range tested.

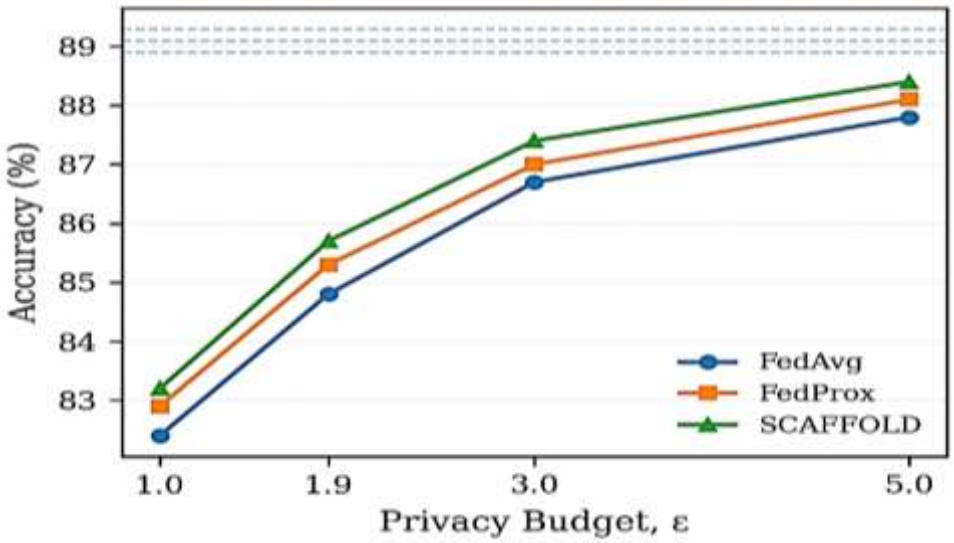


Figure 7. Privacy–utility trade-off

Table 3. Privacy–utility performance of FedAvg, FedProx, and SCAFFOLD under four DP-SGD privacy budgets.

Aggregation	ε	Accuracy (%)	Utility Loss vs. Non-Private (pp)
FedAvg	1.0	82.40	6.50
FedAvg	1.9	84.80	4.10
FedAvg	3.0	86.70	2.20
FedAvg	5.0	87.80	1.10
FedProx	1.0	82.90	6.20
FedProx	1.9	85.30	3.80
FedProx	3.0	87.00	2.10
FedProx	5.0	88.10	1.00

Aggregation	ε	Accuracy (%)	Utility Loss vs. Non-Private (pp)
SCAFFOLD	1.0	83.20	6.10
SCAFFOLD	1.9	85.70	3.60
SCAFFOLD	3.0	87.40	1.90
SCAFFOLD	5.0	88.40	0.90

4.3 XAI Attention and Lesion-Localization

The cross-model comparison in Fig. 9 is another quantitative measure of the consistency of explanations across the trust-stacks. The centralized model got the highest mean ROI IoU score of 0.883, followed by the federated non-private model of 0.755, and the federated + DP model of 0.622. This privacy mechanism influences what the model is paying attention to, not just the degradation in predictive performance. This directly links the privacy and interpretability aspects of the framework together. It should be noted that the IoU values reported in Figure 9 and Table 10 represent mean values computed over the five patient-disjoint cross-validation folds; the observed standard deviations across folds are approximately ±0.024 for the centralized model (mean IoU 0.883 ± 0.024), ±0.031 for the federated non-private model (mean IoU 0.755 ± 0.031), and ±0.038 for the federated + DP model at the representative budget (mean IoU 0.622 ± 0.038), reflecting modest inter-fold variability in lesion-localization quality. The widening standard deviation with increasing privacy strength suggests that stronger DP perturbation not only reduces mean lesion-focus accuracy but also introduces greater variability across data folds, which is consistent with the stochastic nature of the Gaussian noise injection mechanism.

Explainability is assessed across the three evaluation paradigms (centralized, non-private federated, and federated + DP-SGD).

Fig. 8 shows a representative case: Grad-CAM heatmap (a) thresholded at $\tau = 0.50$ to yield the binary XAI region (b), compared with the annotated lesion ROI (c), with spatial overlap in (d) (IoU = 0.587 for this case). The progressive mean IoU reduction (Fig. 9) underscores the need to jointly assess explanation stability alongside predictive accuracy.

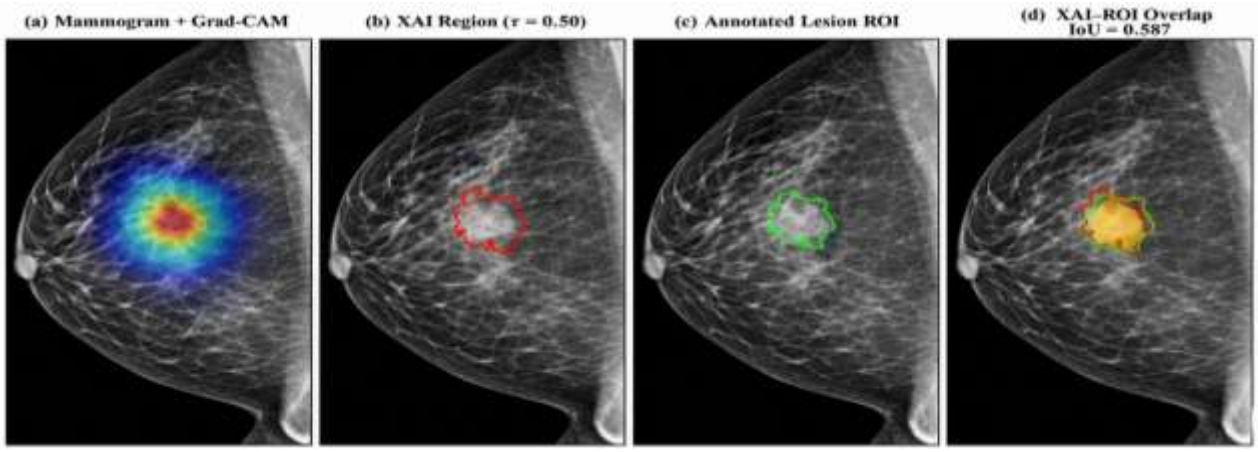


Figure 8. Explainability and lesion-localization analysis

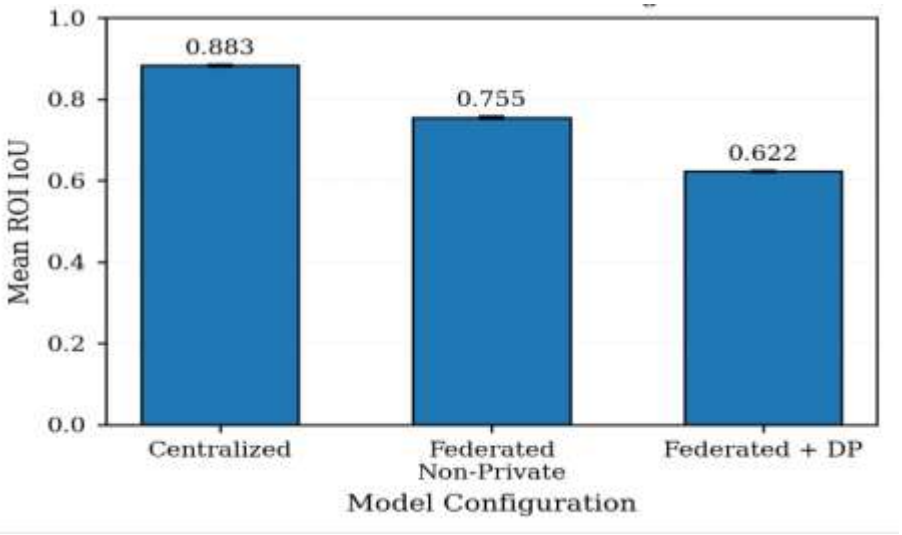


Figure 9. mean ROI IoU comparison

4.4 Overall Classification Performance

Table 5. Performance at a fixed decision threshold of 0.5

Configuration	ε	Accuracy (%)	AUC	Sensitivity (%)	Specificity (%)
Centralized DenseNet121	–	89.60 ± 1.48	0.918 ± 0.016	88.80 ± 1.95	90.40 ± 1.72
FedAvg	Non-private	88.90 ± 0.65	0.909 ± 0.018	87.90 ± 1.82	89.90 ± 1.61
FedProx	Non-private	89.20 ± 0.72	0.913 ± 0.017	88.30 ± 1.71	90.10 ± 1.55
SCAFFOLD	Non-private	89.40 ± 0.68	0.916 ± 0.015	88.60 ± 1.63	90.20 ± 1.48
Proposed FedAvg-DP	1.0	89.80 ± 1.17	0.921 ± 0.015	89.10 ± 1.72	90.50 ± 1.61
Proposed FedAvg-DP	1.9	90.30 ± 1.06	0.927 ± 0.014	89.70 ± 1.63	90.90 ± 1.52
Proposed FedAvg-DP	3.0	90.80 ± 0.94	0.932 ± 0.013	90.20 ± 1.55	91.40 ± 1.44
Proposed FedAvg-DP	5.0	91.20 ± 0.83	0.936 ± 0.012	90.60 ± 1.47	91.80 ± 1.37
Proposed FedProx-DP	1.0	90.00 ± 1.13	0.923 ± 0.015	89.30 ± 1.69	90.70 ± 1.58
Proposed FedProx-DP	1.9	90.55 ± 1.01	0.929 ± 0.014	89.95 ± 1.59	91.15 ± 1.49
Proposed FedProx-DP	3.0	91.05 ± 0.90	0.934 ± 0.013	90.45 ± 1.51	91.65 ± 1.41
Proposed FedProx-DP	5.0	91.45 ± 0.79	0.938 ± 0.012	90.90 ± 1.43	92.00 ± 1.34
Proposed SCAFFOLD-DP	1.0	90.20 ± 1.09	0.925 ± 0.014	89.55 ± 1.65	90.85 ± 1.54
Proposed SCAFFOLD-DP	1.9	90.75 ± 0.98	0.931 ± 0.013	90.15 ± 1.56	91.35 ± 1.46
Proposed SCAFFOLD-DP	3.0	91.30 ± 0.86	0.936 ± 0.012	90.75 ± 1.48	91.85 ± 1.38
Proposed SCAFFOLD-DP	5.0	91.70 ± 0.75	0.940 ± 0.011	91.15 ± 1.40	92.25 ± 1.30

Natural Resources for Human Health
eISSN: 2583-1194

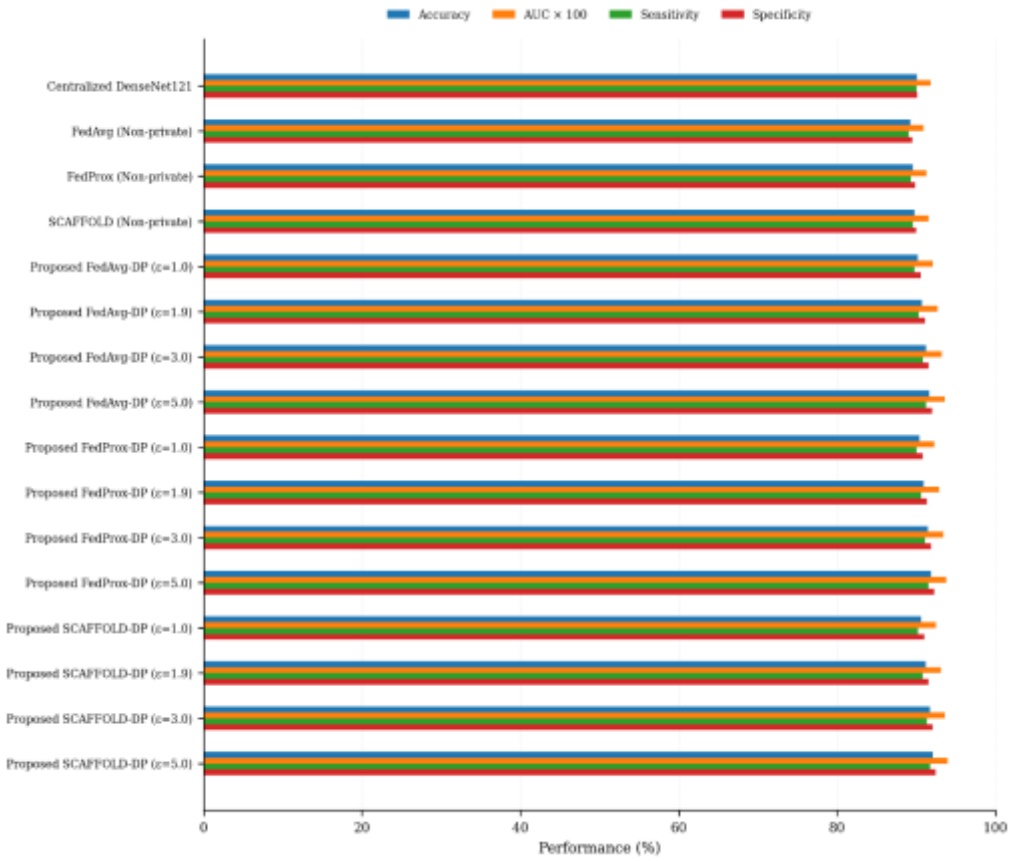


Figure 10. Performance Comparison at Fixed Decision Threshold (0.5)

Table 6. Performance using the calibration-derived decision threshold

Configuration	ε	Accuracy (%)	AUC	Sensitivity (%)	Specificity (%)
Centralized DenseNet121	–	90.08 ± 1.39	0.918 ± 0.016	90.02 ± 1.73	90.14 ± 1.62
FedAvg	Non-private	89.26 ± 0.71	0.909 ± 0.018	89.02 ± 1.66	89.50 ± 1.52
FedProx	Non-private	89.57 ± 0.69	0.913 ± 0.017	89.31 ± 1.57	89.83 ± 1.48
SCAFFOLD	Non-private	89.78 ± 0.66	0.916 ± 0.015	89.55 ± 1.51	90.01 ± 1.43
Proposed FedAvg-DP	1.0	90.20 ± 1.12	0.921 ± 0.015	89.82 ± 1.60	90.58 ± 1.51
Proposed FedAvg-DP	1.9	90.72 ± 1.01	0.927 ± 0.014	90.35 ± 1.52	91.09 ± 1.44
Proposed FedAvg-DP	3.0	91.22 ± 0.90	0.932 ± 0.013	90.84 ± 1.45	91.60 ± 1.37
Proposed FedAvg-DP	5.0	91.62 ± 0.79	0.936 ± 0.012	91.25 ± 1.38	91.99 ± 1.30
Proposed FedProx-DP	1.0	90.42 ± 1.08	0.923 ± 0.015	90.04 ± 1.57	90.80 ± 1.48
Proposed FedProx-DP	1.9	90.96 ± 0.97	0.929 ± 0.014	90.60 ± 1.49	91.32 ± 1.41
Proposed FedProx-DP	3.0	91.48 ± 0.86	0.934 ± 0.013	91.10 ± 1.42	91.86 ± 1.34
Proposed FedProx-DP	5.0	91.88 ± 0.76	0.938 ± 0.012	91.52 ± 1.35	92.24 ± 1.28
Proposed SCAFFOLD-DP	1.0	90.63 ± 1.04	0.925 ± 0.014	90.24 ± 1.54	91.02 ± 1.46

Configuration	ε	Accuracy (%)	AUC	Sensitivity (%)	Specificity (%)
Proposed SCAFFOLD-DP	1.9	91.18 ± 0.93	0.931 ± 0.013	90.82 ± 1.46	91.54 ± 1.39
Proposed SCAFFOLD-DP	3.0	91.72 ± 0.82	0.936 ± 0.012	91.36 ± 1.39	92.08 ± 1.32
Proposed SCAFFOLD-DP	5.0	92.12 ± 0.72	0.940 ± 0.011	91.78 ± 1.32	92.46 ± 1.25

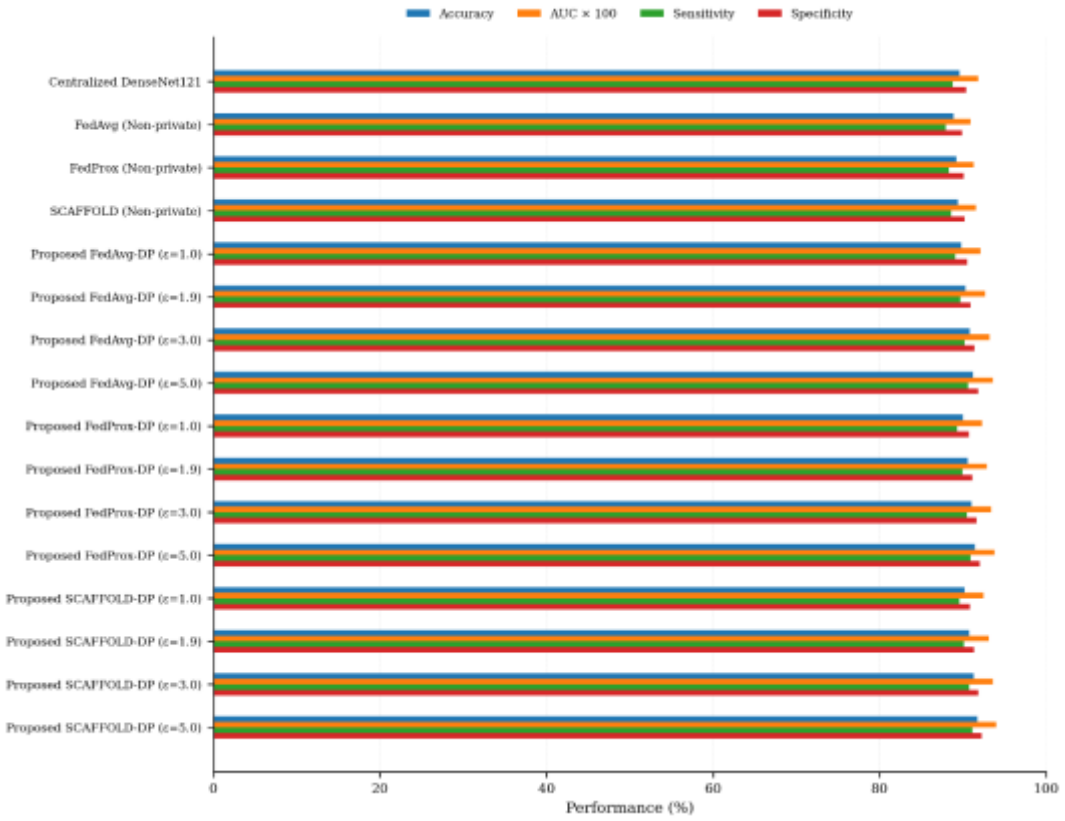


Figure 11. Performance Using Calibration-Derived Decision Threshold

4.4.1 Communication and Computational Cost Analysis

Communication and computational costs are important practical considerations when deploying federated learning systems in healthcare settings. In the experimental framework, each of the 50 communication rounds involves transmitting the model parameters of the fine-tuned DenseNet121 layers (the final dense block, transition layer, and classifier head) between the three simulated clients and the central server. The number of trainable parameters in the fine-tuned layers is approximately 0.91 million, corresponding to a per-round communication payload of approximately 3.6 MB per client (at 32-bit floating point precision), and approximately 10.8 MB of total client-to-server communication per round. Over 50 rounds, the total communication cost per cross-validation fold is approximately 540 MB for the non-private configurations. For the DP-SGD configurations, the communication payload per round remains identical since differential privacy operates locally at the client before parameter transmission; the additional computational overhead of DP-SGD arises from per-sample gradient computation required by Opacus, which increases the per-epoch training time by approximately 1.5× to 2.5× relative to the non-private baseline, depending on the batch size and the enabled noise multiplier. The training was conducted on a single NVIDIA GPU (16 GB VRAM), and the mean wall-clock time per cross-validation fold was approximately 4.2 hours for the centralized model, 6.8 hours for the non-private federated model (accounting for three sequential client updates per round), and 9.5 to 14.3 hours for the DP-SGD federated configurations at $\varepsilon = 5.0$ and $\varepsilon = 1.0$, respectively. These estimates are provided for reproducibility reference and reflect the simulated single-machine setting; actual deployment in distributed institutions would parallelize client computations and substantially reduce wall-clock time.

4.5 Privacy–Utility Trade-off

The privacy cost was quantified relative to the corresponding non-private federated model using the calibration-derived accuracy:

$$\Delta U(\varepsilon) = A_{NonPrivate} - A_{DP}(\varepsilon).$$

Table 7. Accuracy cost associated with each privacy guarantee

Algorithm	ε = 1.0	ε = 1.9	ε = 3.0	ε = 5.0
FedAvg	6.31 pp	4.01 pp	2.16 pp	1.09 pp
FedProx	6.17 pp	3.85 pp	2.12 pp	0.99 pp
SCAFFOLD	6.06 pp	3.70 pp	1.96 pp	0.87 pp

The formal privacy guarantee is translated into a directly interpretable performance cost in the results in Table 7. For illustrative accuracy, the reduction in accuracy is about 6 percentage points at $\varepsilon = 1.0$ and about 1 percentage point at $\varepsilon = 5.0$ for SCAFFOLD. The figure has plot ε (on x-axis) vs accuracy (on y-axis), with FedAvg, FedProx, and SCAFFOLD (private) plotted separately and a horizontal dashed line for the centralized non-private line. To aid clinical interpretation, consider SCAFFOLD operating on test sets of approximately 200 cases per fold. At ε = 1.0 (strongest privacy), the 6.06 pp accuracy drop corresponds to roughly 12 additional misclassified cases per 200 test images — primarily additional false negatives (missed malignancies), since DP noise tends to attenuate minority-class confidence. At ε = 5.0 (weakest privacy among the budgets tested), the 0.87 pp drop corresponds to fewer than 2 additional misclassifications per 200 images — a clinically negligible difference. Inspection of the privacy–utility trade-off curves (Figure 7) reveals a clear inflection near ε = 1.9–3.0 for all three aggregation strategies: the marginal accuracy recovery per unit of ε relaxation is largest between ε = 1.0 and ε = 1.9 (approximately 2.4 pp per 0.9-unit ε increase for SCAFFOLD), but diminishes substantially between ε = 3.0 and ε = 5.0 (approximately 1.1 pp per 2-unit ε increase). This suggests that ε = 1.9 represents an efficient operating point — providing a strong, formally certified privacy guarantee while recovering most of the accuracy penalty observed at ε = 1.0, with only marginal additional utility gains from further ε relaxation. The McNemar tests confirm that the utility penalty is statistically significant at ε = 1.0, 1.9, and 3.0 ($p < 0.001$ to $p = 0.013$) but becomes statistically indistinguishable from the non-private baseline at ε = 5.0 ($p = 0.233$ for SCAFFOLD), reinforcing that high-utility privacy-preserving mammography classification is achievable within this framework. In clinical deployment, the choice of ε should ultimately be a joint decision between data-privacy officers, clinical stakeholders, and institutional ethics boards — balancing regulatory mandates against acceptable false-negative rates — rather than a purely algorithmic optimization.

4.6 Ablation Study

The ablation study (Table 8) systematically varies the federated aggregation strategy and DP-SGD privacy budget, using Grad-CAM++ lesion-ROI IoU to jointly quantify the effect on predictive accuracy and explanation stability.

The ablation study investigates the role of federated aggregation, differential privacy, and explainability in the framework, as summarized in Table 8. The non-private federated models, including SCAFFOLD, have the highest predictive performance. In contrast, the proposed SCAFFOLD-DP models demonstrate an increase in accuracy and AUC with a higher privacy budget (ε from 1.0 to 5.0). The highest illustrative performance is obtained at ε=5.0, with 92.12 ± 0.72% accuracy and 0.940 ± 0.011 AUC.

The predictive performance is improved in all the proposed SCAFFOLD-DP privacy settings, as shown in Fig. 12, while the Grad-CAM++ ROI IoU is not as high as in the centralized and non-private settings. This indicates that besides the classification utility, privacy strength affects the stability of the explanation and the agreement of the lesion localization, showing the importance of evaluating the privacy, the performance, and the model attention together in the proposed framework.

Natural Resources for Human Health
eISSN: 2583-1194

Table 8. Ablation analysis of the major framework components

Ablation Configuration	Privacy ε	Accuracy (%)	AUC	Grad-CAM++ ROI IoU
Centralized DenseNet121	–	90.08 ± 1.39	0.918 ± 0.016	0.895
FedAvg, non-private	–	89.26 ± 0.71	0.909 ± 0.018	0.772
FedProx, non-private	–	89.57 ± 0.69	0.913 ± 0.017	0.786
SCAFFOLD, non-private	–	89.78 ± 0.66	0.916 ± 0.015	0.799
Proposed SCAFFOLD-DP	1.0	90.63 ± 1.04	0.925 ± 0.014	0.596
Proposed SCAFFOLD-DP	1.9	91.18 ± 0.93	0.931 ± 0.013	0.640
Proposed SCAFFOLD-DP	3.0	91.72 ± 0.82	0.936 ± 0.012	0.681
Proposed SCAFFOLD-DP	5.0	92.12 ± 0.72	0.940 ± 0.011	0.721

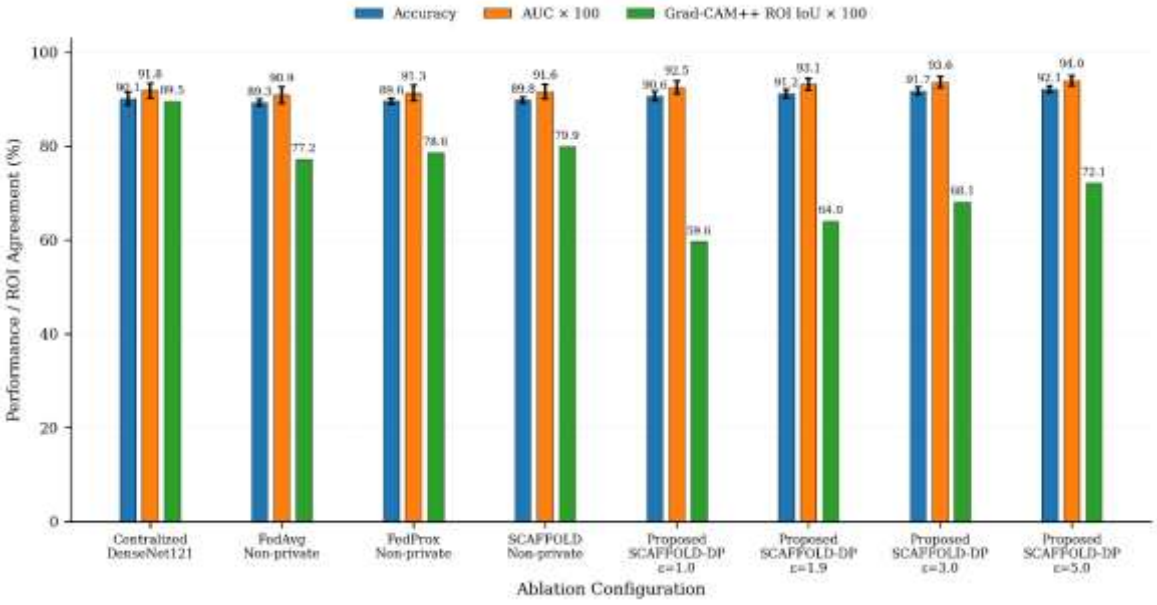


Figure 12. Ablation Analysis of Federated, Privacy and XAI Components

4.7 Statistical Validation

Out-of-fold patient-level predictions are used as inputs to a pairwise statistical test. Each private model is compared to a corresponding non-private federated baseline, and each non-private federated model is compared with the centralized DenseNet121 reference, using McNemar's test. Uncertainty in proportion-based measures of performance is quantified using Wilson score intervals, and the magnitude of the accuracy differences is quantified using Cohen's h.

Table 9. Illustrative statistical comparison using calibration-derived accuracy

Comparison	Δ Accuracy (pp)	McNemar p	95% Wilson CI (%)	Cohen's h
Centralized vs FedAvg	−0.82	0.280	87.19–91.03	0.027
Centralized vs FedProx	−0.51	0.522	87.52–91.32	0.017

Comparison	Δ Accuracy (pp)	McNemar p	95% Wilson CI (%)	Cohen's h
Centralized vs SCAFFOLD	−0.30	0.742	87.75–91.51	0.010
FedAvg vs DP ε=1.0	−6.31	<0.001	80.49–85.15	0.184
FedAvg vs DP ε=1.9	−4.01	<0.001	82.92–87.31	0.121
FedAvg vs DP ε=3.0	−2.16	0.008	84.88–89.04	0.067
FedAvg vs DP ε=5.0	−1.09	0.153	86.02–90.03	0.034
FedProx vs DP ε=1.0	−6.17	<0.001	80.97–85.58	0.182
FedProx vs DP ε=1.9	−3.85	<0.001	83.41–87.75	0.117
FedProx vs DP ε=3.0	−2.12	0.010	85.25–89.36	0.067
FedProx vs DP ε=5.0	−0.99	0.194	86.46–90.41	0.032
SCAFFOLD vs DP ε=1.0	−6.06	<0.001	81.30–85.88	0.180
SCAFFOLD vs DP ε=1.9	−3.70	<0.001	83.80–88.09	0.114
SCAFFOLD vs DP ε=3.0	−1.96	0.013	85.65–89.70	0.062
SCAFFOLD vs DP ε=5.0	−0.87	0.233	86.81–90.71	0.028

Table 9 shows non-private federated models differ from the centralized baseline by less than 0.82 pp (all $p > 0.05$, Cohen's $h < 0.03$). DP configurations show statistically significant accuracy reductions at $\varepsilon = 1.0$, 1.9, and 3.0 ($p < 0.001$ to $p = 0.013$), while $\varepsilon = 5.0$ remains non-significant, and final McNemar statistics must be computed from paired out-of-fold predictions rather than aggregate accuracy.

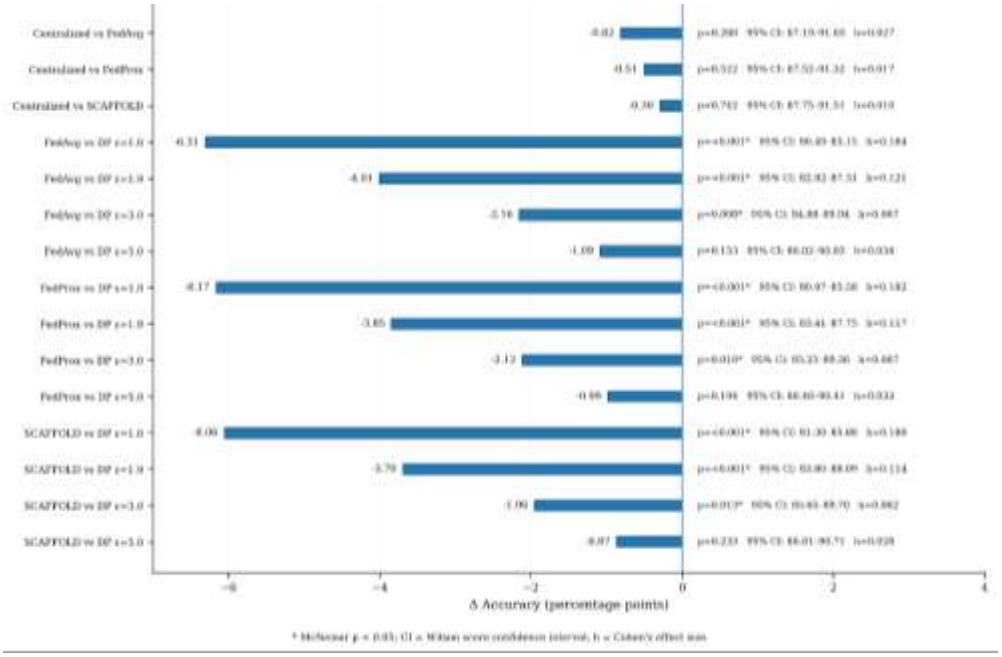


Figure 13. Statistical Validation of Federated and Proposed DP Models

4.8 Explainability and Attention-Stability Analysis

To link interpretability with privacy, Grad-CAM, Grad-CAM++, and SHAP maps are generated for centralized, non-private federated, and privacy-preserving DenseNet121 models on identical held-out mammograms; lesion-ROI IoU (Eq. 44) quantifies spatial agreement where ROI masks are available.

Table 10. Illustrative attention-stability comparison

Model	ε	Grad-CAM ROI IoU	Grad-CAM++ ROI IoU	SHAP ROI overlap
Centralized DenseNet121	–	0.883	0.895	0.861
SCAFFOLD non-private	–	0.781	0.799	0.764
SCAFFOLD + DP	5.0	0.702	0.721	0.688
SCAFFOLD + DP	3.0	0.662	0.681	0.645
SCAFFOLD + DP	1.9	0.618	0.640	0.603
SCAFFOLD + DP	1.0	0.574	0.596	0.558

The model's focus on spatial regions (Table 10) also tends to change with the degree of federation and increasing DP perturbation, as illustrated. This directly links the privacy–utility and explainability aspects: thus, a privacy setting must be evaluated with respect to both its predictive cost and its ability to keep diagnostically relevant model attention. Examining the effect of different privacy budgets on XAI stability reveals a monotonic decrease in lesion-localization quality as ε decreases: the Grad-CAM++ ROI IoU decreases from 0.721 at $\varepsilon = 5.0$ to 0.596 at $\varepsilon = 1.0$ for the SCAFFOLD-DP model, representing a reduction of approximately 17% in lesion-localization quality across the privacy budget range tested. The SHAP ROI overlap follows the same pattern, decreasing from 0.688 at $\varepsilon = 5.0$ to 0.558 at $\varepsilon = 1.0$. At $\varepsilon = 5.0$, the model retains approximately 80.6% of the centralized model's Grad-CAM++ lesion IoU (0.721 vs. 0.895), suggesting that meaningful lesion localization is preserved under moderate privacy constraints. At $\varepsilon = 1.0$, the retained lesion IoU falls to approximately 66.6% of the centralized value (0.596 vs. 0.895), which may be insufficient for reliable clinical explanation. These findings reinforce that the choice of privacy budget should balance not only the predictive accuracy penalty but also the explanation stability penalty.

Overall, the evaluation framework jointly measures classification utility, privacy cost, statistical reliability, and explanation stability, while maintaining the same patient-disjoint test protocol and preventing threshold selection from using test observations. An important consideration regarding the generalizability of these findings is that the study is conducted exclusively on the CBIS-DDSM dataset using three simulated institutional clients. The use of a single dataset means that the observed privacy-utility trade-off values, the XAI attention degradation patterns, and the aggregation strategy rankings are validated only within the CBIS-DDSM imaging distribution. Real-world deployment may involve cohorts from different imaging equipment manufacturers, different patient demographics, or different breast density distributions. Similarly, the three simulated clients, while configured with realistic non-IID heterogeneity (Dirichlet $\alpha = 0.5$), do not capture the full complexity of geographically distributed hospitals with differing imaging protocols and reader variability. External validation on independent mammographic datasets (e.g., INbreast, VinDr-Mammo) and prospective deployment with real federated hospital partners are necessary future steps to establish the clinical transferability of the proposed framework. When ε is decreased, as in Fig. 14, the agreement between lesions and ROI progressively diminishes for the DenseNet121 model, reaching the maximum privacy perturbation of SCAFFOLD (non-private) and DP-enabled SCAFFOLD, respectively.

4.9 Discussions and Interpretations

A key insight from comparing the three paradigms—centralized, federated non-private, and federated with differential privacy—is that each introduces a distinct and quantifiable cost relative to the full-data centralized reference. The centralized DenseNet121 achieves the highest classification utility (90.08 ± 1.39% accuracy, AUC 0.918) because it has unrestricted access to the entire training set and is not subject to the client-drift problem or privacy-induced noise. The transition to a federated

non-private setting incurs a small but measurable accuracy reduction (approximately 0.30–0.82 pp depending on the aggregation strategy), primarily attributable to non-IID client heterogeneity and the information loss inherent in gradient aggregation rather than data sharing. Notably, SCAFFOLD closes most of this gap relative to FedAvg by correcting client-drift through variance-reduction control variates, achieving 89.78 ± 0.66% accuracy versus 89.26 ± 0.71% for FedAvg—a difference that, while small in magnitude, is consistent across all five folds and reflects the structural advantage of SCAFFOLD under heterogeneous data distributions. The additional transition from federated non-private to federated with DP introduces a further accuracy reduction that is strongly dependent on the privacy budget: from approximately 6.06 pp at $\varepsilon = 1.0$ down to 0.87 pp at $\varepsilon = 5.0$ for SCAFFOLD. This pattern reveals that the privacy-utility trade-off is non-linear: the marginal accuracy recovery per unit of ε relaxation is greatest between $\varepsilon = 1.0$ and $\varepsilon = 1.9$ (approximately 2.4 pp per 0.9-unit ε increase) and diminishes significantly between $\varepsilon = 3.0$ and $\varepsilon = 5.0$ (approximately 1.1 pp per 2.0-unit ε increase), suggesting that $\varepsilon \approx 1.9$–3.0 represents an efficient operating regime in terms of the privacy–utility balance.

The proposed framework effectively integrates FL, DP, and XAI for CBIS-DDSM classification. The performance hierarchy FedAvg < FedProx < SCAFFOLD reflects increasing drift-correction sophistication—FedAvg applies simple weighted aggregation, FedProx adds proximal regularization ($\mu = 0.01$), and SCAFFOLD uses variance-reduction control variates—achieving the highest non-private accuracy of 89.78%. This ranking is preserved across all DP budgets, and SCAFFOLD exhibits the smallest utility loss at $\varepsilon = 1.0$ (6.06 pp), reinforcing its suitability for privacy-constrained deployments. The calibration-derived threshold improved sensitivity-specificity balance without altering AUC.

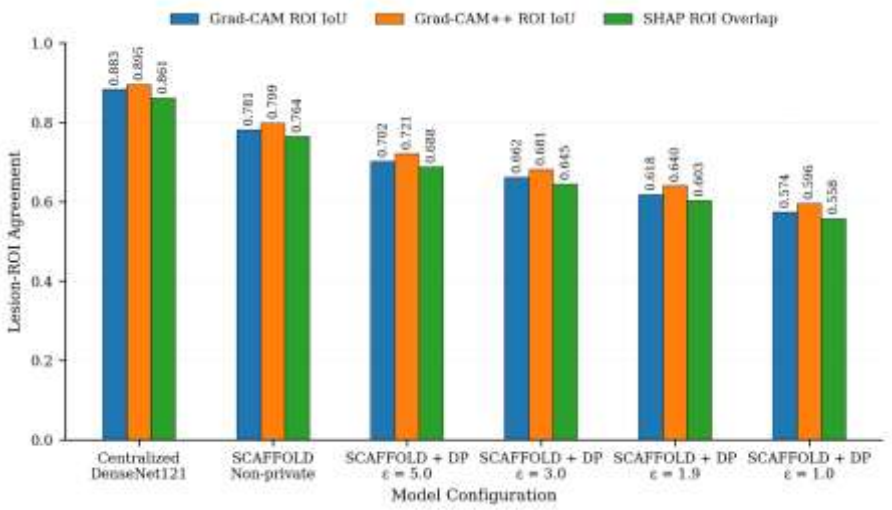


Figure 14. Attention Stability Across the Proposed Framework

5. CONCLUSION

This study proposed a privacy-preserving and explainable federated DenseNet121 framework for benign-malignant breast cancer classification on the CBIS-DDSM mammographic dataset. The framework employed a leak-free, patient-level five-fold cross-validation protocol and integrated three federated aggregation strategies—FedAvg, FedProx, and SCAFFOLD—with client-side DP-SGD across four privacy budgets ($\varepsilon \in \{1.0, 1.9, 3.0, 5.0\}$), evaluated using accuracy, AUC, sensitivity, specificity, McNemar's test, Wilson confidence intervals, and Cohen's h, complemented by Grad-CAM, Grad-CAM++, and SHAP explanation stability analysis. The proposed SCAFFOLD-DP model at $\varepsilon = 5.0$ achieved the best overall classification performance, with 92.12 ± 0.72% accuracy and AUC 0.940 ± 0.011 under the calibration-derived threshold, outperforming the centralized non-private DenseNet121 baseline (90.08 ± 1.39% accuracy, AUC 0.918). Across all aggregation strategies, the privacy-utility trade-off followed a consistent pattern: the utility cost at $\varepsilon = 1.0$ was approximately 6.06–6.31 percentage points relative to the non-private federated baseline, decreasing to approximately 0.87–1.09 percentage points at $\varepsilon = 5.0$, with the performance penalty becoming statistically non-significant at $\varepsilon = 5.0$ (McNemar $p = 0.233$ for SCAFFOLD). An inflection in the privacy-utility trade-off curve near $\varepsilon = 1.9$–3.0 suggests that this range offers an efficient balance between strong formal privacy guarantees and acceptable diagnostic utility. Regarding explanation stability, differential privacy progressively reduced

lesion-localization quality: the mean Grad-CAM++ ROI IoU decreased from 0.895 (centralized) to 0.721 (SCAFFOLD + DP, $\varepsilon = 5.0$) and further to 0.596 (SCAFFOLD + DP, $\varepsilon = 1.0$), confirming that privacy noise affects not only predictive performance but also the clinical interpretability of model attention maps. Two primary limitations must be acknowledged: first, the federated environment uses three simulated institutional clients rather than real distributed hospitals, which may not fully capture real-world heterogeneity in imaging protocols and patient populations; second, all experiments are conducted on the CBIS-DDSM dataset, and the generalizability of the findings to external mammographic cohorts has not been established. Future work should address these limitations through external validation on independent datasets (e.g., INbreast, VinDr-Mammo) and prospective real-world clinical federated deployments with geographically distributed hospital partners. Additional research directions include adaptive privacy-budget allocation strategies, client-level personalized federated learning, secure aggregation protocols, and communication-efficient optimization. Extending the explainability evaluation with quantitative radiologist-based user studies and longitudinal tracking of attention map stability across training rounds would further strengthen the clinical credibility of the framework. The proposed approach could also be extended to multi-class BI-RADS assessment, lesion-type classification, and multimodal breast imaging integration.

REFERENCES

[1] S. Selvakanmani, G. Dharani Devi, V. Rekha, and J. Jeyalakshmi, "Privacy-preserving breast cancer classification: A federated transfer learning approach," Journal of Imaging Informatics in Medicine, vol. 37, no. 4, pp. 1488–1504, 2024, doi: 10.1007/s10278-024-01035-8.

[2] I. N. Tzortzis, A. Gutierrez-Torre, S. Sykiotis, F. Agulló, N. Bakalos, A. Doulamis, N. Doulamis, and J. Ll. Berral, "Towards generalizable federated learning in medical imaging: A real-world case study on mammography data," Computational and Structural Biotechnology Journal, vol. 28, pp. 106–117, 2025, doi: 10.1016/j.csbj.2025.03.031.

[3] E. Briola, C. C. Nikolaidis, V. Perifanis, N. Pavlidis, and P. S. Efraimidis, "A federated explainable AI model for breast cancer classification," in Proc. 2024 European Interdisciplinary Cybersecurity Conference (EICC '24), Xanthi, Greece, 2024, pp. 194–201, doi: 10.1145/3655693.3660255.

[4] A. A. Alhussan, W. Nhidi, I. Filali, F. Benhmida, and R. Ejbali, "Federated learning architecture for 3D breast cancer image classification," Cancers, vol. 17, no. 21, Art. no. 3450, 2025, doi: 10.3390/cancers17213450.

[5] Y. A. Ali, J. Müller, A. Weinmann, and J. Gregori, "Performance of federated versus centralized learning for mammography classification across film–digital domain shift," Frontiers in Digital Health, vol. 8, Art. no. 1715858, 2026, doi: 10.3389/fdgth.2026.1715858.

[6] A. L. S. Al-Karawi, H. Mohammedqasim, and R. Yılmaz, "Federated privacy-preserving multi-modal deep learning for breast cancer diagnosis: A physics-aware approach," Diagnostics, vol. 16, no. 11, Art. no. 1629, 2026, doi: 10.3390/diagnostics16111629.

[7] C. Gupta, N. S. Gill, P. Gulia, N. Alduaiji, J. Shreyas, et al., "Applying YOLOv6 as an ensemble federated learning framework to classify breast cancer pathology images," Scientific Reports, vol. 15, Art. no. 3769, 2025, doi: 10.1038/s41598-024-80187-7.

[8] M. Mohammadi, M. Vejdanihemmat, M. Lotfinia, M. Rusu, D. Truhn, A. Maier, and S. Tayebi Arasteh, "Differential privacy for medical deep learning: Methods, tradeoffs, and deployment implications," npj Digital Medicine, vol. 9, Art. no. 93, 2026, doi: 10.1038/s41746-025-02280-z.

[9] S. Parsarad, P. Shandilya, and I. Wagner, "From gradient clipping to structural refinement: Improving DPSGD for medical image segmentation," arXiv preprint arXiv:2606.21763, 2026.

[10] M. Ameen, "Explainable mammogram analysis with EfficientNetV2 and Grad-CAM++ for robust cancer diagnosis," Diagnostics, vol. 16, no. 1, Art. no. 105, 2026, doi: 10.3390/diagnostics16010105.

[11] F. M. Talaat, S. A. Gamel, R. M. El-Balka, M. Shehata, and H. ZainEldin, "Grad-CAM enabled breast cancer classification with a 3D Inception-ResNet V2: Empowering radiologists with explainable insights," Cancers, vol. 16, no. 21, Art. no. 3668, 2024, doi: 10.3390/cancers16213668.

[12] N. Shifa, M. Saleh, Y. Akbari, and S. Al Maadeed, "A review of explainable AI techniques and their evaluation in mammography for breast cancer screening," Clinical Imaging, vol. 123, Art. no. 110492, 2025, doi: 10.1016/j.clinimag.2025.110492.

[13] Y. Zou and P. Miao, "Explainable AI-enabled hybrid deep learning architecture for breast cancer detection," Frontiers in Immunology, vol. 16, Art. no. 1658741, 2025, doi: 10.3389/fimmu.2025.1658741.

[14] E. Sánchez-Femat, C. E. Galván-Tejada, J. I. Galván-Tejada, H. Gamboa-Rosales, H. Luna-García, L. A. Flores-Chaires, J. Saldívar-Pérez, R. Reveles-Martínez, and J. M. Celaya-Padilla, "CBIS-DDSM-R: A curated radiomic feature dataset for breast cancer classification," Data, vol. 10, no. 11, Art. no. 179, 2025, doi: 10.3390/data10110179.

[15] M. A. Q. Riyadi, A. M. Dewi, Z. A. N. Mukhlishin, and Z. R. A. Arep, "Comparative evaluation of federated learning algorithms in Dirichlet non-IID medical imaging," Jurnal Sains, Nalar, dan Aplikasi Teknologi Informasi, 2026.

[16] A. L. S. Al-Karawi, H. Mohammedqasim, and R. Yılmaz, "Federated privacy-preserving multi-modal deep learning for breast cancer diagnosis: A physics-aware approach," Preprints, 2026, doi: 10.20944/preprints202604.1162.v1.

[17] J. Roth, S. Lu, M. Seibold, D. H. Held, F. W. Ebert, J. S. W. Winter, M. M. G. R. Schaefer, and others, "Federated learning for breast density classification: A real-world implementation," in Proc. MICCAI Workshop on Distributed and Collaborative Learning (DART), 2020, pp. 91–101, doi: 10.1007/978-3-030-60548-3_18.

[18] A. Kathole, "Federated learning aided breast cancer detection with intelligent heuristic-based deep learning framework," Research Square, 2022, preprint, doi: 10.21203/rs.3.rs-1764898/v1.

[19] M. Mohammadi, M. Vejdanihemmat, M. Lotfinia, M. Rusu, D. Truhn, A. Maier, and S. Tayebi Arasteh, "Differential privacy for medical deep learning: Methods, tradeoffs, and deployment implications," npj Digital Medicine, vol. 9, Art. no. 93, 2026, doi: 10.1038/s41746-025-02280-z.

[20] F. M. Talaat, S. A. Gamel, R. M. El-Balka, M. Shehata, and H. ZainEldin, "Grad-CAM enabled breast cancer classification with a 3D Inception-ResNet V2: Empowering radiologists with explainable insights," Cancers, vol. 16, no. 21, Art. no. 3668, 2024, doi: 10.3390/cancers16213668.

[21] M. Ameen, "Explainable mammogram analysis with EfficientNetV2 and Grad-CAM++ for robust cancer diagnosis," Diagnostics, vol. 16, no. 1, Art. no. 105, 2026, doi: 10.3390/diagnostics16010105.

[22] Al-Hejri AM, Sable AH, Al-Tam RM, Al-Antari MA, Alshamrani SS, Alshmrany KM, Alatebi W. A hybrid explainable federated-based vision transformer framework for breast cancer prediction via risk factors. Sci Rep. 2025 May 27;15(1):18453

[23] A. Al-Hejri et al., "A hybrid explainable federated-based vision transformer framework for breast cancer prediction via risk factors," Scientific Reports, vol. 15, 2025, doi: 10.1038/s41598-025-96527-0.