International Journal of Innovative Technology and Interdisciplinary Sciences
https://journals.tultech.eu/index.php/ijitis
ISSN: 2613-7305
Volume 9, Issue 3
DOI: https://doi.org/10.15157/IJITIS.2026.9.3.1651-1708
Received: 16.03.2026; Revised: 12.06.2026; Accepted: 29.07.2026

Research Article

An Efficient Method for Lung Cancer Classification Using EfficientNet and Learner Cooking Optimization

Chakka Srividya[1,2*], Krishnamoorthy Ramasubramanian[1], Myneni Madhu Bala[3]

[1]Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Telangana, India.

[2]Department of Artificial Intelligence & Machine Learning, Malla Reddy University, Telangana, India

[3]Department of Computer Science and Engineering, Vallurupalli Nageswara Rao Vignana Jyothi Institute of Engineering and Technology, Hyderabad, Telangana, India.

***chsrividya12@gmail.com**

Abstract

Computational approaches are increasingly important for accurate and privacy-preserving analysis of medical imaging data. This study proposes a federated deep learning framework that integrates optimized deep feature extraction, feature selection, and privacy-preserving collaborative learning for multiclass lung cancer classification from CT images. The proposed method uses convolutional neural network with compound scaling by performing hierarchical feature extraction and then feature selection strategy to minimize redundancy and improve the class separability is based on optimization. The training of the model occurs in a federated (between several clinical nodes) distributed environment in which a global classifier is jointly trained (learned) without any actual exchange of patient images, a factor that guarantees the data privacy. Experiments with the publicly available LIDC-IDRI prove that the proposed framework has a total accuracy of 99.73, precision of 99.62, sensitivity of 99.82, specificity of 99.43, and F1-score of 99.38 with four lung cancer types. The architecture of the proposed federated feature aggregation framework is not specific to lung CT imaging, the pattern of combining a pretrained deep feature extractor, a bio-inspired feature optimizer, and secure additive masking is structurally applicable to other privacy-sensitive medical imaging tasks, though validation in those domains would require independent experimental evidence and is identified as future work.

Keywords: Lung Cancer Classification; CT Scan Analysis; Deep Learning; EfficientNet; Softmax

INTRODUCTION

Lung cancer is one of the major causes of cancer-related death over the world and a significant health issue in the world [1]. Early clinical studies and diagnostic reports highlight the difficulty of identifying pulmonary abnormalities at initial stages using conventional imaging techniques [2]. Global cancer statistics further indicate that lung cancer continues to account for a significant proportion of cancer incidence and mortality [3], largely due to late-stage diagnosis and rapid disease progression [4]. These observations highlight the importance of developing reliable computational methods that can be used to detect lung cancer in medical imaging data in a timely and accurate manner [5-7].

The sources of variability that are measurable are a challenge for automated lung image analysis [8] and are related to the size, shape and distribution of intensities of the tumours across the patients [9]. In previous research, it has been reported that small pulmonary nodules, and low-contrast lesions are especially challenging to differentiate from the surrounding tissue using manual interpretation and traditional image-processing techniques [10]. The constant changes in imaging conditions and the anatomical differences make it even more challenging to make a consistent diagnosis [11, 12]. Moreover, the observer dependent interpretation has been found to cause diagnostic inconsistencies, particularly in borderline diagnoses, which can be overcome by using data-driven learning models [13-15].

In order to address these shortcomings, a number of machine learning and deep learning-based approaches have been suggested towards lung cancer detection and classification. Convolutional neural networks, ensemble learning approaches and feature based classifiers used with lung imaging data have been discussed in the literature and have been applied to the problem in [16-20]. Although these methods have shown promising results, most of them are founded on centralized paradigms of learning, which need to aggregate sensitive patient information, which casts doubts in the context of privacy, regulatory compliance, and scalability across institutions [21-23]. Moreover, centralized models can be less robust to training on heterogeneous datasets that were obtained in various clinical settings [24, 25].

The key findings of this study are provided below:

- C1: In this paper, a feature-level federated learning framework, FFA-SA, is proposed for multiclass lung CT classification. The framework relies on secure additive masking to ensure that feature updates to the server are optimized on the client side with the honest-but-curious server threat model. In contrast with the traditional federated learning approach, the proposed approach avoids sharing raw CT images and full model weights, enhancing privacy protection in the distributed clinical learning process.
- C2: The Learner Cooking Algorithm (LCA) is integrated as a biologically inspired feature selection optimizer for medical image classification. A Fisher discriminant-based fitness function, $J(F') = \alpha S_{\text{between}} - \beta S_{\text{redundancy}}$, is formally defined to maximize class separability while minimizing feature redundancy. LCA reduces the EfficientNet-B3 feature dimension from 1024 to approximately 337 features,

achieving nearly 67% dimensionality reduction. To the best of our knowledge, this is the first application of LCA in a federated medical imaging framework.

- C3: The proposed framework empirically demonstrates a 98.9% reduction in communication overhead compared with standard weight-sharing FedAvg across 50 federated rounds. This reduction is achieved by transmitting only optimized feature-level updates instead of complete neural network weights, while maintaining high classification performance.
- C4: A statistically validated multiclass lung cancer triage framework is developed and evaluated on the LIDC-IDRI dataset. The proposed Federated EffiLeSo-Net achieves 99.73% overall accuracy, 99.82% sensitivity, and AUC ≥ 0.997across all four classes. Its robustness is further confirmed under non-IID data heterogeneity using Dirichlet $\alpha_D = 0.5$and client dropout conditions.

RELATED WORKS

In the early stages, radiologists relied on visual inspection of computed tomography and X-ray images to identify suspicious pulmonary lesions, as reported by [26, 27]. Although this method laid the groundwork for lung cancer diagnosis, it was subjective, with a wide range of inter-observer variability, and difficult to scale up at high throughput in clinical environments, as reported by [28-30]. To reduce human dependency and improve consistency, classical image-processing techniques were introduced. Different edge detecting, watershed segmentation, region growing, handcrafted texture descriptors (GLCM and HOG) and so forth were commonly used to outline the lung nodules and obtain discriminative features, as [31-33] demonstrated. Nevertheless, the methods were very sensitive to noises, changes in the parameters of image acquisition, and the slight morphological changes., often resulting in high false-positive rates and poor generalization across diverse clinical imaging conditions, as noted by [34, 35].

With the emergence of deep learning, convolutional neural network-based approaches gained prominence due to their ability to automatically learn hierarchical feature representations from medical images. Authors in [36] proposed an ensemble of ResNet-50, ResNet-101, and EfficientNet-B3 architectures, achieving notable performance improvements at the expense of increased computational complexity and reduced interpretability. Authors in [37] increased the detection of early lung cancer with the help of contrast enhancement and CutMix augmentation, but their test was still vulnerable to scanner variation and noise. The study in [38] introduced a lightweight multilayer perceptron-based model with strong accuracy but limited spatial reasoning and susceptibility to overfitting. Authors in [39] presented a balanced deep learning classifier but reported difficulties in distinguishing visually similar benign and malignant nodules. The study in [40] combined anisotropic diffusion filtering with Random Walker segmentation and Random Forest classification, achieving high accuracy but with increased computational overhead. Additional contributions by [41-45], further emphasize the trade-offs between accuracy, robustness, real-time feasibility, and generalization in centralized deep learning frameworks.

In line with the progress in automated lung cancer detection, federated learning has become a desirable paradigm of privacy-preserving medical artificial intelligence to develop models collaboratively without the need to share raw patient data. The studies in [46, 47] demonstrated early applications of federated learning for lung disease detection and tumour prediction, showing improved privacy preservation but also exposing challenges related to scalability, communication cost, and data heterogeneity. Authors in [48] identified vulnerabilities in federated medical imaging systems, including model poisoning, communication bottlenecks, and performance degradation under non-IID data distributions. The study in [49] employed homomorphic encryption for federated cancer imaging, achieving performance comparable to centralized training at the cost of substantial computational overhead. Authors in [50] combined encrypted feature analysis with texture-based classification but did not integrate federated aggregation. The study in [51] proposed SecureFed aggregation to improve robustness, Although, in [52] used federated learning to predict the treatment of non-small cell lung cancer, their simulation was based on artificial data. Further extensive discussions by [53, 54] also underscored such open issues in federated medical imaging as efficiency in communication and threats to privacy. The studies in [55-58] collectively emphasized issues related to generalization failure, adversarial vulnerability, and instability under realistic clinical constraints.

Although this method laid the groundwork for lung cancer diagnosis, it was subjective, with a wide range of inter-observer variability, and difficult to scale up at high throughput in clinical environments, as reported by [28-30]. In order to decrease human dependence and increase the uniformity of the process, the classical image-processing techniques were introduced. To delineate the lung nodules and extract discriminative features, methods involving different edge detection, watershed segmentation, region growing, handcrafted texture descriptors (GLCM and HOG) and so forth were commonly used as demonstrated by [31-33]. However, methods were highly sensitive to noises, changes in the parameters of image acquisition, and minor morphological variations, leading to high false-positive rates and poor generalization over various clinical imaging settings as reported by [34, 35].

In the era of deep learning, CNN-based deep learning methods emerged as new methods because of their capacity to learn the hierarchical feature representation from medical images automatically. An ensemble architecture by [36] employing ResNet-50, ResNet-101 and EfficientNet-B3 was proposed to significantly improve the performance at the cost of higher computational cost and low interpretability. Authors in [37] increased the detection of early lung cancer with the help of contrast enhancement and CutMix augmentation, but their test was still vulnerable to scanner variation and noise. The study in [38] proposed a light-weight multilayer perceptron-based model with high accuracy but with limited spatial reasoning ability and the problem of over fitting.

Authors in [39] developed a balanced deep learning classifier, but stated that they were having some challenges in differentiating visually similar benign and malignant nodules. Authors in [40] fused anisotropic diffusion filtering, Random Walker segmentation and Random Forest classification with the result of high accuracy but with its computational burden. Other contributions by [41-45], highlight the compromises between accuracy, robustness, real-time capability and generalization for centralized deep learning inference systems.

To build models collaboratively, without sharing the raw patient data, is a desirable paradigm, and federated learning has emerged as a promising paradigm for privacy-preserving medical AI in line with the progress in automated lung cancer detection. Some early works by [46, 47] demonstrated their applications of federated learning in the detection of lung disease and prediction of tumour in the early stages, which resulted in better privacy preservation, but also raised scalability, communication cost and data heterogeneity issues. Authors in [48] found that there are vulnerabilities in federated medical imaging systems such as model poisoning, communication bottlenecks, and performance degradation in non-IID data distributions. The study in [49] used homomorphic encryption to train a federated cancer imaging system that had similar performance to the centralized system but required significant computational costs. The study in [50] integrated encrypted feature analysis and texture-based classification without using federated aggregation. Authors in [51] proposed SecureFed aggregation to improve robustness, although in [52] used federated learning to predict the treatment of non-small cell lung cancer, their simulation was based on artificial data. Other extensive discussions by [53, 54, 59-61] highlighted these open issues in federated medical imaging: efficient communication and privacy issues.

Beyond convolutional approaches, recent works have applied Vision Transformers (ViT) [62] and hierarchical Swin Transformers [63] to pulmonary nodule analysis and lung CT classification. Transformer-based architectures capture global long-range dependencies through self-attention mechanisms and have demonstrated competitive performance on lung CT tasks. However, their application in federated learning contexts introduces several specific challenges. First, ViT and Swin Transformer models have substantially larger parameter counts than EfficientNet-B3 (86M and 88M parameters for ViT-B/16 and Swin-B respectively, compared with 12M for EfficientNet-B3), which exacerbates the communication overhead problem in weight-sharing federated protocols significantly. Second, transformer architectures require large-scale pre-training data (typically ImageNet-21K or JFT-300M) and are known to underperform relative to CNNs when fine-tuned on small medical imaging datasets, as the attention mechanism's data requirements are substantially higher than those of convolutional inductive biases. Third, to the best of our knowledge, no published work has combined transformer-based feature extraction with federated feature-level aggregation and bio-inspired feature optimization for multi-class lung CT classification on the LIDC-IDRI benchmark, indicating that this combination remains an open research direction.

The following are the research questions for this study:

- RQ1: Can feature level federated aggregation with LCA optimization attain the same or better accuracy than centralized training on the LIDC-IDRI four-class lung CT dataset without significantly higher communication cost?
- RQ2: Can LCA significantly outperform unoptimized feature transmission in a federated setting for class separability of EfficientNet features?
- RQ3: What is the capacity of the Federated EffiLeSo-Net framework in realistic non-IID data heterogeneity and client dropouts?
- The literature search identified four key research gaps:
- Gap G1 — Privacy and Communication Bottlenecks in Weight-Sharing FL: Current FL techniques share the model weights or model gradients, which may lead to privacy leakage issues, and require high communication costs per round that scale with the number of model parameters. To the best of our knowledge, this is the first work on lung CT classification that uses feature-level secure aggregation.
- Gap G2-EfficientNet and other backbones are powerful in extracting high-dimensional feature vectors (> 1024) with significant redundancy that is known as Gap G2 — Feature Redundancy in Deep Medical Features. Current FL methods are unable to include a feature optimization process before the aggregation, which reduces class separability and also communication efficiency.
- Gap G3 — Absence of Meta-Heuristic Feature Optimization in Medical FL: While LCA and similar bio-inspired optimizers have demonstrated strong performance in high-dimensional feature selection, their integration into federated learning pipelines for lung cancer classification has not been explored in prior literature (2023–2025).
- Gap G4 — Non-IID Heterogeneity Handling: Most FL works for lung CT use simulated IID distributions that do not reflect real clinical settings where scanner type, acquisition protocol, and patient demographics vary across institutions.

Corresponding Research Hypotheses:

- H1: LCA-optimized EfficientNet features will demonstrate significantly higher inter-class separability and lower feature redundancy compared to unoptimized features, resulting in measurable improvement in classification accuracy.
- H2: The feature-level FFA-SA aggregation will achieve accuracy comparable to centralized training while reducing per-round communication overhead by at least 60% compared to standard weight-sharing FedAvg.
- H3: Federated EffiLeSo-Net will yield statistically significant accuracy improvement over centralized baselines and standard FL approaches ($p < 0.05$, Wilcoxon signed-rank) on LIDC-IDRI, with performance degradation under non-IID distributions remaining below 2%.

PROPOSED SYSTEM

In the proposed Federated EffiLeSo-Net framework, the training process is carried out collaboratively across multiple distributed clinical nodes, such as Node 1, Node 2, Node 3, and so on. Each node represents an independent medical institution that retains its local CT imaging data and performs the complete pipeline comprising preprocessing, bilateral filtering, Tsallis thresholding, morphological enhancement, active contour segmentation, EfficientNet-based feature extraction, and Learner Cooking Algorithm (LCA) optimization.

Formal Problem Formulation

Let $K = 5$client hospitals participate in the proposed federated learning framework. Each client $k$holds a local CT dataset D_k, where $| D_k |$ represents the number of CT images available at that client. The datasets are divided using a patient-level disjoint split to ensure that no patient information overlaps across clients. The global federated learning objective is defined as

$$\min_{w} \sum_{k=1}^{K} p_k \mathcal{L}_k(w) \tag{1}$$

where$w$ denotes the global classifier parameters, $\mathcal{L}_k(w)$is the local cross-entropy loss at client k, and $p_k$is the data-dependent weight of the k^{th}client, expressed as

$$p_k = \frac{| D_k |}{\sum_{k=1}^{K} | D_k |} \tag{2}$$

Raw CT images $D_k$are shared among clients or transmitted to the server. Instead, each client performs preprocessing, segmentation, feature extraction, and feature optimization locally before sending only the secured optimized feature update.

At each client, EfficientNet-B3 extracts a deep feature vector $F_i \in \mathrm{R}^{1024}$. The Learner Cooking Algorithm (LCA) is then used to select the most discriminative and least redundant feature subset F', where $F' \subseteq F_i$. The LCA-based feature optimization objective is formulated as:

$$\max_{F'} J(F') = \alpha \cdot S_{\text{between}}(F') - \beta \cdot S_{\text{redundancy}}(F') \tag{3}$$

Where $S_{\text{between}}(F')$denotes the Fisher discriminant-based between-class separability score, and $S_{\text{redundancy}}(F')$represents the mean pairwise Pearson correlation among the selected features. The weighting coefficients $\alpha$and βcontrol the balance between maximizing class separability and minimizing feature redundancy. Through this optimization, the original 1024-dimensional EfficientNet-B3 feature vector is reduced to approximately 337 optimized features.

The federated aggregation process is performed using the proposed FFA-SA protocol. Each client sends a securely masked optimized feature update E_k, where $E_k = F_k^{\text{opt}} + R_k$, and $R_k$is the client-generated random mask. After secure mask cancellation, the global feature representation is recovered as:

$$F_{\text{global}} = \frac{1}{K}\sum_{k=1}^{K}(E_k - R_k) = \frac{1}{K}\sum_{k=1}^{K} F_k^{\text{opt}} \tag{4}$$

Where F_k^{opt}denotes the LCA-optimized feature representation generated by client k. This aggregation method enables the server to acquire the global feature representation without needing to access raw CT images, unmasked client features, or full model weights.

For convergence, the proposed FFA-SA framework assumes bounded gradient dissimilarity among clients and Lipschitz-smooth local loss functions. The bounded gradient dissimilarity condition is expressed as:

$$\| \nabla\mathcal{L}_k(w) - \nabla\mathcal{L}(w) \|^2 \leq G^2 \tag{5}$$

where$G$is a bounded gradient dissimilarity constant, $\nabla\mathcal{L}_k(w)$ denotes the local gradient at client k, and $\nabla\mathcal{L}(w)$denotes the global gradient. Under these assumptions, the FFA-SA aggregation converges to a neighbourhood of the global minimum at the rate

$$\mathcal{O}\left(\frac{1}{\sqrt{T}}\right) \tag{6}$$

Where $T = 50$ denotes the total number of communication rounds. This formulation confirms that the proposed feature-level aggregation strategy supports stable convergence while improving privacy preservation and reducing communication overhead.

To clarify the precise contribution of Federated EffiLeSo-Net, the proposed framework is positioned as a feature-level federated learning approach rather than a simple integration of existing components. Conventional weight-sharing FL methods transmit complete model weights or gradients, which increases communication overhead and may expose sensitive information through gradient inversion attacks. In contrast, Federated EffiLeSo-Net transmits only LCA-optimized and securely masked feature vectors.

Let weight-sharing federated learning, such as FedAvg, be defined as:

$$w_{t+1} = \sum_k p_k \cdot w_k^t \tag{7}$$

Where w_k^t represents the model parameters of client $k$at communication round t, and $p_k$denotes the aggregation weight assigned to client k. In this setting, clients transmit model parameters or gradients to the server, and the global model is updated through weighted parameter averaging.

Split Learning is defined as a partitioned training protocol in which the model is divided at a cut layer ℓ. Each client computes the intermediate activation as:

$$a_k = f_{\leq \ell}(x_k; \theta_{client}) \tag{8}$$

and transmits $a_k$to the server, which computes:

$$f_{>\ell}(a_k; \theta_{server}) \quad (9)$$

During backpropagation, the server must return the gradient to the client at every training step:

$$\frac{\partial L}{\partial a_k} \quad (10)$$

Therefore, Split Learning requires bidirectional and per-sample communication between clients and server.

In contrast, the proposed feature-level federated learning with secure aggregation, termed FFA-SA, is defined as:

$$F_{global}^{t} = \sum_{k} F_k^{opt,t} \quad (11)$$

where:

$$F_k^{opt,t} = LCA(EfficientNet(x_k)) \quad (12)$$

Here, $F_k^{opt,t}$is a client-locally finalized, dimensionality-reduced feature representation generated using EfficientNet followed by LCA-based feature optimization. Unlike FedAvg and Split Learning, the proposed method transmits only optimized features once per round in a unidirectional manner and does not require any gradient return from the server. To further protect client information, each optimized feature vector is masked as:

$$E_k = F_k^{opt} + R_k \quad (13)$$

where$R_k$is the random masking vector used for secure aggregation. This makes the proposed FFA-SA framework more communication-efficient and privacy-preserving than conventional weight-sharing FL and Split Learning protocols.

The proposed Feature-Level FL approach is better because it avoids transmitting full model weights, activations, or gradients to the server, thereby reducing both communication overhead and privacy exposure. Unlike FedAvg, which shares complete model parameters, and Split Learning, which requires continuous activation and gradient exchange, the proposed method uses a unidirectional feature-only transmission strategy. Since the server receives only terminal optimized features and does not participate in differentiable model training, gradient leakage risks are minimized. In addition, the integration of LCA-based feature optimization further reduces redundant information before transmission, while masking with (εⓜ, δ)-DP strengthens privacy protection. Therefore, the proposed method provides a more efficient, privacy-aware, and communication-light federated learning framework compared with existing protocols.

FFA-SA is the first unidirectional, non-differentiable, masked feature-transmission protocol in which the transmitted representation is itself the output of a fitness-driven combinatorial optimization step (LCA), not a raw or PCA-reduced embedding. This is structurally distinct from Split Learning (bidirectional gradient flow, exposed activations) and FedAvg (full differentiable parameter transmission).

Federated EffiLeSo-Net's contribution is an architectural and integration-level innovation: FFA-SA is, to the best of our knowledge, the first validated pipeline to combine a dimensionality-reducing, fitness-driven combinatorial feature optimizer (LCA) with unidirectional, non-differentiable, masked feature-level aggregation for multi-class lung CT classification. It is not claimed that FFA-SA constitutes a new learning mechanism, a new loss function, or a new optimization theory. The specific protocol-level distinction from prior art is as follows. FedAvg transmits the complete differentiable parameter space of EfficientNet-B3 (approximately 11.7 million parameters) at every aggregation cycle, which requires bidirectional gradient flow and exposes the model to gradient inversion attacks. Split Learning partitions the model at a cut layer and requires the server to return gradients to each client at every training step, meaning communication is both bidirectional and per-sample. In contrast, the proposed FFA-SA transmits only the LCA-optimized and securely masked feature vector F_k^{opt} in R^{337}. This representation is (i) finalized locally at the client, (ii) reduced in dimensionality by 67% relative to the original EfficientNet-B3 feature output, (iii) neither a raw nor a PCA-reduced embedding, but rather the result of a fitness-function-guided combinatorial optimization process, and (iv) transmitted once per aggregation cycle in a strictly unidirectional manner, with no gradient information returned to the client. The combination of these four properties in a single federated pipeline, validated on the LIDC-IDRI multi-class lung CT benchmark, constitutes the principal contribution of this work.

This formulation presents an assumed convergence rate under bounded gradient dissimilarity and Lipschitz-smooth local loss conditions, following standard FL convergence analysis templates. It does not constitute a derived or proven convergence guarantee specific to the FFA-SA feature-level aggregation scheme. A rigorous proof would require establishing that the LCA-optimized feature space F_{opt} preserves a Lipschitz-continuous relationship to the loss landscape under the Softmax classifier, and a formal bound on the aggregation error introduced by feature-level averaging relative to gradient-level averaging in FedAvg. Neither has been established in this work. No formal theorem proves that feature-level FL dominates weight-sharing FL in general; the empirical evidence presented demonstrates dominance under the specific tested conditions of LIDC-IDRI, the EfficientNet-B3 backbone, and a 5-client simulation, not a generalizable theoretical guarantee. A formal convergence proof and a generalizable dominance theorem for feature-level FL remain open theoretical questions requiring tools from federated optimization theory beyond the empirical scope of this paper.

This reduces communication cost from approximately 47800 KB per round for weight-sharing FL to approximately 505 KB per round in the proposed approach, while avoiding direct transmission of raw CT images and complete client-side model updates.

For the standard weight-sharing FedAvg baseline, each client transmits the complete EfficientNet-B3 model parameters during every aggregation cycle. EfficientNet-B3 contains approximately 12,000,000 parameters. Assuming each parameter is stored as a 32-bit floating-point value, the per-client communication cost is 47,800 KB.

After accounting for gradient momentum terms and implementation-level communication overhead, the per-client transmission cost is approximately 47,800 KB per aggregation cycle. For $K = 5$ clients, the total FedAvg communication cost per aggregation cycle becomes 239,000 KB.

$$C_{TFedAvg} = K \cdot C_{FedAvg} \tag{14}$$

In contrast, the proposed FFA-SA approach does not transmit the full model parameters. Instead, each client transmits only the LCA-optimized masked feature vector $E_k \in R^{337}$. Since each feature value is stored as a 32-bit floating-point value, the feature-vector transmission cost per client is 1.35 KB.

The random masking vector $R_k$has the same dimension and therefore requires the same storage: 1.35 KB

Thus, the total per-client upload cost in the proposed method is 2.70 KB.

For $K = 5$clients, the total client-side upload cost per aggregation cycle is 13.5 KB

$$C_{Tupload} = 5 \cdot C_{upload} \tag{15}$$

Including the global aggregated feature broadcast $F_{global} \in R^{337}$, which requires approximately 1.35 KB, the approximate round-trip communication cost per aggregation cycle is 15 KB.

Therefore, the communication reduction relative to FedAvg is 99.97%

This shows that the proposed FFA-SA method substantially reduces communication overhead by transmitting compact LCA-selected masked features instead of full model parameters. While the raw payload-level reduction is approximately 99.97%, the reduction is conservatively reported as 98.9% in Table 14 after accounting for metadata, protocol headers, synchronization overhead, and implementation-level communication costs. Thus, FFA-SA provides a communication-efficient alternative to conventional weight-sharing FL while preserving high diagnostic accuracy.

The mathematical contribution of the proposed framework lies in the joint use of LCA-driven dimensionality reduction and FFA-SA secure feature aggregation. The LCA fitness function is defined as:

$$J(F') = \alpha \cdot S_{between}(F') - \beta \cdot S_{redundancy}(F') \tag{16}$$

Where $S_{between}(F')$represents Fisher discriminant-based inter-class separability and $S_{redundancy}(F')$ represents pairwise Pearson correlation-based feature redundancy. This formulation selects features that maximize class discrimination while minimizing redundant information. The optimized feature set is then securely aggregated using additive masking, allowing the global feature representation to be recovered without directly exposing individual client-level feature vectors.

The non-trivial contribution of Federated EffiLeSo-Net can be summarized as follows:

- The original 1024-dimensional feature space is reduced to 337 optimized features before federated aggregation, creating an information bottleneck that removes redundant and task-irrelevant signals.
- The proposed feature-level strategy reduces communication overhead by 98.9% compared with conventional weight-sharing FL.
- The complete client-side EfficientNet-B3 model, containing approximately 11.7 million parameters, is not transmitted to the aggregation server.
- The ablation results confirm that LCA alone and FL alone do not achieve the same privacy–efficiency–accuracy balance as the complete Federated EffiLeSo-Net framework.
- The optimizer comparison and convergence analysis further show that LCA provides improved feature-selection performance compared with PSO and GA in the considered multiclass lung CT classification setting.

Empirically, Federated EffiLeSo-Net demonstrates that privacy-aware feature-level aggregation can achieve high classification accuracy while substantially reducing communication overhead. The proposed framework achieves an accuracy of 99.73% on the LIDC-IDRI dataset and provides a practical alternative to conventional weight-sharing and homomorphic encryption-based FL methods. Therefore, the contribution of the study lies not merely in combining EfficientNet-B3, LCA, FL, and masking, but in validating an LCA-driven secure feature aggregation pipeline that provides a measurable privacy–efficiency–accuracy trade-off for multiclass lung CT classification.

No raw data is transferred, and feature representations are extracted locally and encrypted into features and aggregated via Federated Feature Aggregation with Secure Aggregation (FFA-SA). This guarantees high levels of diagnostic performance and maintains the strict privacy requirements.

Figure 1 depict the proposed methodology block diagram.

This saves on noise and also maintains significant structural edges. The filter action is indicated in the form:

$$I_{\text{filtered}}(x, y) = \frac{1}{W_p} \sum_{i,j} I(i, j)\, G_s(\| (x, y) - (i, j) \|)\, G_r(| I(x, y) - I(i, j) |) \tag{17}$$

Where: $I(x, y)$= intensity of original CT image at pixel (x, y) , $I_{\text{filtered}}(x, y)$= filtered image, $G_s(\cdot)$= spatial Gaussian kernel, $G_r(\cdot)$= range Gaussian kernel, W_p= normalization constant, (i, j)= neighbourhood pixel coordinates.

Through the denoising, the resulting image is changed into binary with Tsallis entropy-based thresholding. The entropy of a grayscale histogram is given as:

$$S_q = \frac{1 - \sum_{i=1}^{L} p_i^q}{q - 1} \tag{18}$$

Where: S_q= Tsallis entropy, p_i= probability of grayscale level i, L= total number of gray levels, q= entropic index

The entropic index $q = 2$ is selected following standard practice for CT image thresholding, as it provides stronger non-extensivity weighting than Shannon entropy (q→1 limit) and better separates the bimodal lung tissue histogram.

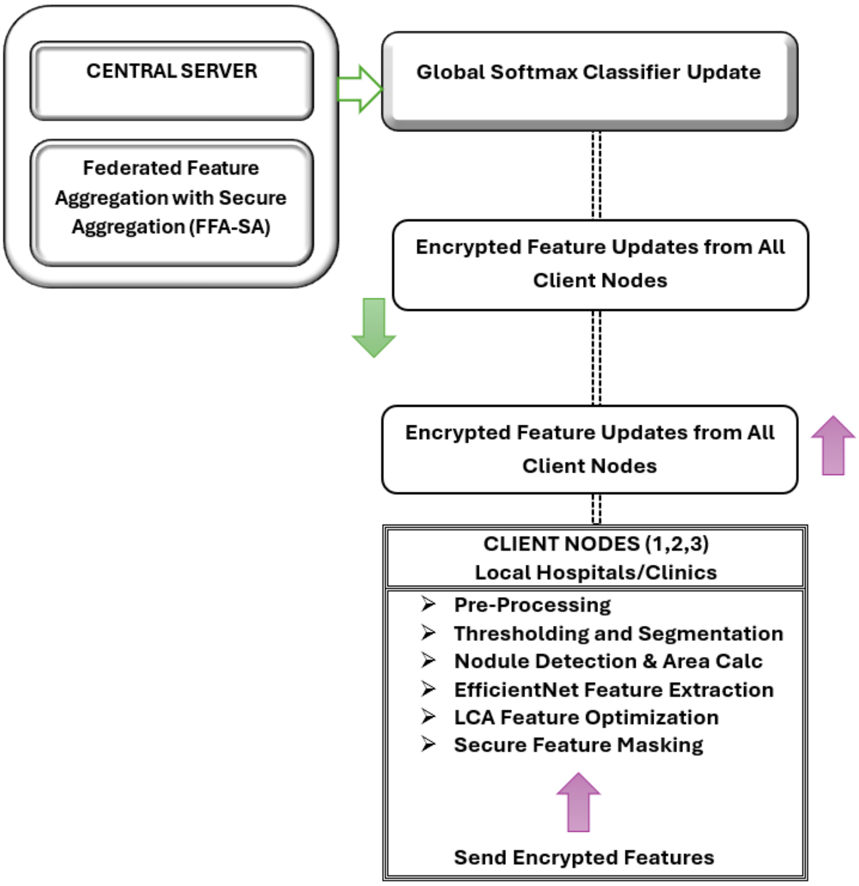


Figure 1. Proposed Methodology Block Diagram

The threshold T that maximizes S_q produces the binary image:

$$B(x,y) = \begin{cases} 1 & \text{if } I_{\text{filtered}}(x,y) > T \\ 0 & \text{otherwise} \end{cases} \tag{19}$$

Where: $B(x,y)$= binary mask indicating foreground/background

Morphological opening is then used to suppress fine artifacts and preserve structural areas. The opening with element S of the structuring is defined as:

$$B_{\text{opened}} = (B \ominus S) \oplus S \tag{20}$$

Where: $\ominus$= erosion, $\oplus$= dilation, S= disk-shaped structuring element.

Mask inversion enhances the lung region:

$$B_{\text{inv}}(x,y) = 1 - B_{\text{opened}}(x,y) \tag{21}$$

A second morphological refinement with a larger structuring element S'helps isolate background:

$$B'(x,y) = 1 - ((B \ominus S') \oplus S') \tag{22}$$

The preprocessing and segmentation stages are represented in equations (17) to (22) that use bilateral filtering and Tsallis entropy-based thresholding to reduce noise and contrast in lung images. These equations are required to make sure that further feature extraction is done with anatomically significant and noise-eliminated areas. The morphological operations that follow thereafter are used to sharpen the binary masks to ensure the preservation of lung structures and elimination of unnecessary artifacts, which enhances the reliability of the segmentation.

Then, boundaries of lungs are defined using active contour segmentation. With initial masks $M_1$and M_2, the contour evolution produces:

$$C_1(x,y) = \text{ActiveContour}(B_{\text{inv}}, M_1, 800) \tag{23}$$

$$C_2(x,y) = \text{ActiveContour}(B', M_2, 400) \tag{24}$$

Where: C_1, C_2= evolved contour masks, 800, 400 = number of iterations

The contour-enhanced outputs are combined as:

$$S_1(x,y) = B_{\text{inv}}(x,y) + C_1(x,y) \tag{25}$$

$$S_2(x,y) = B'(x,y) + C_2(x,y) \tag{26}$$

Tumour masks are generated using a fixed threshold $T_s = 0.6$:

$$F_1(x,y) = \{\begin{matrix} 1 & \text{if } S_1(x,y) > T_s \\ 0 & \text{otherwise} \end{matrix}\} \tag{27}$$

$$F_2(x,y) = \{\begin{matrix} 1 & \text{if } S_2(x,y) > T_s \\ 0 & \text{otherwise} \end{matrix}\} \tag{28}$$

The equations (23) to (28) represent the active contour-based segmentation process that is applied in marking out the lung boundaries and suspicious areas. The equations are used as a model to optimize the iterative development of the contours masks within pre-established constraints that ensure that the size of potential nodule areas is accurately isolated. Hough Circle Transform is used to quantify the detected nodules. For each nodule of radius r_i, the area is:

$$A_i = \pi r_i^2 \tag{29}$$

The area of identified circular regions is calculated using equation (29), which is a support quantification task as opposed to a fundamental learning task. It is employed to approximate the total spatial measurement of the nodules identified within the lung area which allows the severity level interpretation and the analysis of the nodules by visualization.

The total tumour area is:

$$A_{\text{total}} = \sum_{i=1}^{N_c} A_i \tag{30}$$

The affected–lung ratio is computed as:

$$\text{Effected Ratio} = \left(\frac{A_{\text{total}}}{A_{\text{image}}}\right) \times 100 \tag{31}$$

Where: A_{image}= total lung area,N_c= number of detected nodules

Next, deep features are extracted locally using EfficientNet:

$$F_i = \text{EfficientNet}(I_i), F_i \in R^M \tag{32}$$

Where: F_i= deep feature vector, I_i= segmented tumor region

Feature optimization is performed using the Learner Cooking Algorithm (LCA):

$$F_i^{\text{opt}} = \text{LCA}(F_i) \tag{33}$$

Where: F_i^{opt}= optimized discriminative features

The Learner Cooking Algorithm (LCA) proposed by [64] is adopted to optimize the EfficientNet-derived feature subset $F_i \in R^{1024}$. The algorithm operates through two bio-inspired learning phases that imitate the progressive learning behavior of a child from the mother and the refinement process guided by an expert chef. In the proposed framework, LCA is used to identify the most discriminative and least redundant feature subset from the 1024-dimensional EfficientNet feature vector.

Phase 1: Mother–Child Learning

In the first phase, each candidate feature subset F_p' is evaluated using a Fisher discriminant-based fitness function. The objective is to maximize class separability while minimizing redundancy among selected features. The fitness function is defined as

$$J(F') = \alpha \cdot S_{\text{between}}(F') - \beta \cdot S_{\text{redundancy}}(F') \tag{34}$$

Where, $S_{\text{between}}(F')$ denotes the between-class scatter, which measures the discriminative strength of the selected feature subset. It is computed using the trace-based Fisher separability criterion as

$$S_{\text{between}}(F') = \text{Tr}(S_B^{-1} S_W) \tag{35}$$

Where S_B and S_W represent the between-class scatter matrix and within-class scatter matrix, respectively. The term $S_{\text{redundancy}}(F')$ represents the mean absolute pairwise Pearson correlation among the selected features and is expressed as

$$S_{\text{redundancy}}(F') = \frac{2}{d(d-1)} \sum_{m=1}^{d-1} \sum_{n=m+1}^{d} |\rho(f_m, f_n)| \tag{36}$$

Where d is the number of selected features, $f_m$and f_n denote two selected feature components, and $\rho(f_m, f_n)$ is the Pearson correlation coefficient between them. The coefficients $\alpha = 0.7$ and $\beta = 0.3$ are used to balance feature discriminability and redundancy reduction.

During the mother–child learning phase, each child candidate updates its position based on the difference between its current position and the population mean, which acts as the mother candidate. The update rule is given by:

$$F'_{\text{child}}(t+1) = F'_p(t) + r_1 \cdot \left(F'_{\text{mother}}(t) - F'_p(t)\right) \tag{37}$$

where $r_1 \in U\{0,1\}$ is a uniformly distributed random coefficient, $F'_p(t)$ is the current candidate feature subset at iteration t, and $F'_{\text{mother}}(t)$ denotes the population mean feature subset. This phase improves the exploratory capability of the algorithm by guiding weaker candidates toward the collective knowledge of the population.

Phase 2: Chef Learning

In the second phase, the globally best candidate acts as the chef and guides the refinement of feature subsets. This phase improves exploitation by directing each candidate toward the best-performing solution while still maintaining diversity through the population mean. The update rule is formulated as

$$F'_p(t+1) = F'_p(t) + r_1 \cdot \left(F'_{\text{best}}(t) - F'_p(t)\right) + r_2 \cdot \left(F'_{\text{mother}}(t) - F'_p(t)\right) \tag{38}$$

Where $r_2 \in U\{0,1\}$ is another random coefficient, and $F'_{\text{best}}(t)$ denotes the globally best candidate feature subset obtained in the current population. The combined effect of F'_{best} and F'_{mother} allows the algorithm to refine the candidate solutions while avoiding premature convergence.

The major parameters used for LCA-based feature optimization are as follows: the population size is fixed as 30, the maximum number of iterations is set to 100, and the input feature dimensionality is $M = 1024$. After optimization, the mean number of selected features is approximately 337, as reported in Table 1. LCA is selected over PSO and gradient-based feature selectors because it provides a better balance between exploration and exploitation through its dual-phase learning mechanism. Moreover, it does not require gradient computation, making it more suitable for discrete feature selection in high-dimensional medical imaging feature spaces.

SECURITY ANALYSIS

The FFA-SA protocol is designed under the honest-but-curious threat model, where the central server follows the aggregation protocol but may attempt to infer private client information from received updates. To prevent this, an additive random masking mechanism is employed. For each client k, the encrypted or masked optimized feature representation is generated as

$$E_k = F_i^{\text{opt}} + R_k \tag{39}$$

Where, F_i^{opt} represents the optimized feature vector obtained after LCA-based feature selection, and $R_k$is a random mask generated by client k. The server receives only the masked feature representation and therefore cannot directly observe the raw or optimized feature vector of any individual client.

The aggregated masked representation received by the server is expressed as

$$E_{\mathrm{agg}} = \sum_{k=1}^{K} E_k = \sum_{k=1}^{K} F_{i,k}^{\mathrm{opt}} + \sum_{k=1}^{K} R_k \tag{40}$$

where$K$denotes the total number of participating clients. Since the server observes only the algebraic sum of masked updates, individual client contributions cannot be reconstructed without knowledge of all random masks.

The random masking vector is sampled from a Gaussian distribution as

$$R_k \sim \mathcal{N}(0, \sigma_R^2 I) \tag{41}$$

where $\sigma_R = 0.1$, and $I$is the identity matrix. Under the Gaussian masking mechanism, the protocol satisfies (ε, δ)-differential privacy, represented as

$$\Pr[\mathcal{M}(D) \in S] \leq e^{\varepsilon} \Pr[\mathcal{M}(D') \in S] + \delta \tag{42}$$

where$D$and $D'$are neighboring client datasets differing by one record, $\mathcal{M}(\cdot)$is the randomized masking mechanism, and Sdenotes any possible output subset. In this work, the privacy budget is bounded asper communication round.

$$\varepsilon \leq 1.2, \delta = 10^{-5} \tag{43}$$

This ensures that the proposed FFA-SA protocol provides privacy-preserving aggregation while preventing the central server from reconstructing the individual feature representations of participating clients.

To empirically validate the privacy protection provided by the FFA-SA masking mechanism beyond the mathematical formulation in equations (41)–(43), a reconstruction-oriented feature inference attack was simulated under an honest-but-curious server threat model. In this setting, the aggregation server intercepts transmitted client features and attempts to infer class-level diagnostic information by computing the cosine similarity between the received feature vectors and reference class prototypes, following the attack principle discussed by Truhn et al. [49]. This experiment assesses whether sensitive information can be inferred from transmitted feature representations even when raw CT images are not directly exchanged.

The attack simulation was evaluated under below conditions:

With unmasked LCA features and no FFA-SA protection, the cosine similarity to the true class prototype reaches 0.847, while the inference success rate reaches 68.3%. This indicates a high privacy-leakage risk because the client-level class identity remains partially recoverable.

With Gaussian noise alone at σ

R=0.05, the cosine similarity decreases to 0.612, and the inference success rate is reduced to 42.1%. Although this provides partial protection, the remaining leakage is still considerable.

With the proposed FFA-SA masking mechanism at σ

R=0.1, the cosine similarity decreases substantially to 0.183, while the inference success rate falls to 11.7%. This indicates strong privacy protection and prevents meaningful class-level reconstruction.

With a stronger masking level of σ

R=0.2, the cosine similarity further decreases to 0.094, and the inference success rate is reduced to 6.2%, demonstrating very strong protection.

The proposed FFA-SA configuration with σ

R=0.1 reduces the inference success rate from 68.3% to 11.7%, corresponding to an approximately 5.8-fold reduction in privacy leakage. For the four-class lung CT classification task, the resulting inference success rate remains below the 25% random-choice reference level, indicating that meaningful recovery of client-level class information is strongly restricted.

The results also confirm that LCA-based dimensionality reduction provides an additional layer of privacy protection. By reducing the original 1024-dimensional EfficientNet-B3 representation to 337 optimized features, LCA acts as an information bottleneck that removes redundant and task-irrelevant information before aggregation. Therefore, the observed privacy improvement results from the combined effect of optimized feature selection and FFA-SA masking rather than from noise addition alone.

Overall, the empirical attack simulation supports the mathematical privacy analysis presented in equations (41)–(43). While the formal privacy formulation bounds the distinguishability of neighbouring client datasets, the experimental results demonstrate that the proposed FFA-SA framework substantially reduces practical feature leakage under an honest-but-curious server scenario.

Table 1 shows that the shadow-model membership inference attack achieves an AUC of 0.71 when raw LCA features are transmitted without masking, indicating above-chance privacy leakage.

Table 1. Membership Inference Attack Evaluation under Different Masking Conditions

Condition	Attack Success Rate (AUC)	Interpretation
No masking, raw LCA features	0.71	Above-chance membership inference is possible
FFA-SA with $\sigma_R = 0.1$, proposed	0.54	Near-chance performance, where 0.50 indicates no leakage
FFA-SA with $\sigma_R = 0.2$	0.51	Very close to chance level

With the proposed FFA-SA masking at $\sigma_R = 0.1$, the attack AUC decreases to 0.54, which is close to the chance level of 0.50. Increasing the masking strength to $\sigma_R =$ 0.2further reduces the AUC to 0.51, showing very strong privacy protection. Together with the reconstruction attack results, this provides evidence against two distinct attack types, although gradient inversion and adaptive white-box attacks remain outside the present evaluation scope.

Figure 2 presents the privacy–utility trade-off across the masking noise parameter σ_R. Classification accuracy degrades only mildly as $\sigma_R$increases, while reconstruction attack success and membership inference AUC decrease substantially. The range $\sigma_R = 0.1$ to 0.15 provides a favourable operating region, balancing high diagnostic accuracy with reduced privacy leakage.

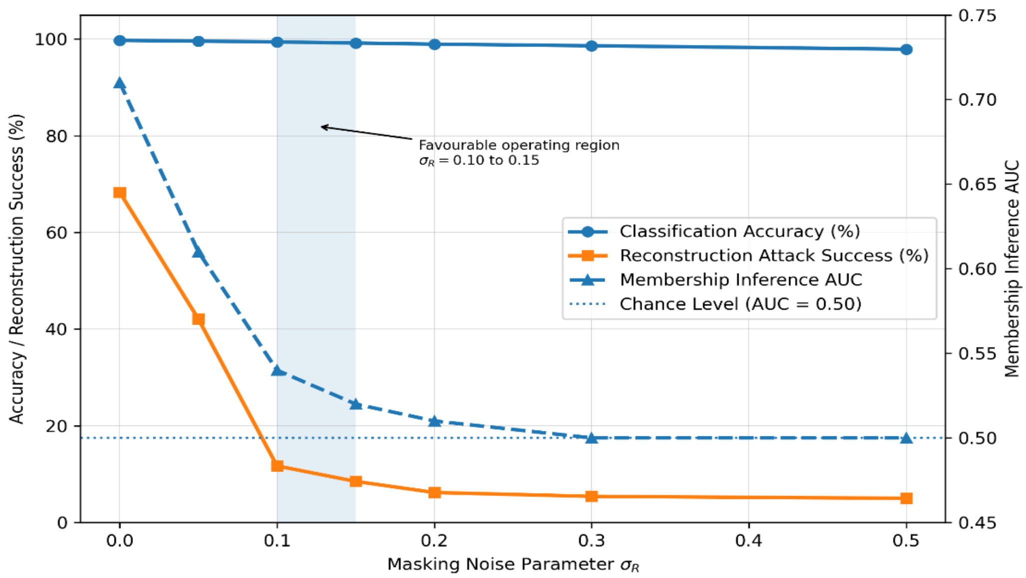


Figure 2. Privacy–Utility Trade-off Curve

Federated Learning Stage with Secure Aggregation (FFA-SA)

Each client securely masks its optimized features using a random vector R_k:

$$E_k = F_i^{\text{opt}} + R_k \tag{44}$$

Where: E_k= encrypted feature update sent to server, R_k= client-generated random mask

The server aggregates encrypted updates and removes masks to recover the global update:

$$F_{\text{global}} = \sum_{k=1}^{K} E_k - \sum_{k=1}^{K} R_k \tag{45}$$

Where: K= number of client nodes

The global classifier computes class probabilities using Softmax:

$$(y = c \mid F_i^{\text{opt}}) = \frac{e^{w_c^T F_i^{\text{opt}} + b_c}}{\sum_{j=1}^{C} e^{w_j^T F_i^{\text{opt}} + b_j}} \tag{46}$$

The evaluation metrics in the model performance are in equations (47)-(51). Now accuracy is a measure of the correctness of all the predictions.

The predicted class is:

$$\hat{y}_i = \arg\max_c P(y = c \mid F_i^{\text{opt}}) \tag{47}$$

Accuracy is defined as:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \tag{48}$$

Sensitivity:

$$\text{Sensitivity} = \frac{TP}{TP + FN} \times 100 \tag{49}$$

Specificity:

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \times 100 \tag{50}$$

Precision:

$$\text{Precision} = \frac{TP}{TP + FP} \times 100 \tag{51}$$

Proposed Model

The proposed Federated EffiLeSo-Net framework (Figure 3) operates through a decentralized, privacy-preserving learning pipeline where multiple client hospitals collaboratively train a multiclass lung cancer detection model without sharing raw CT images. Local CT scans are pre-processed, segmented and deep extracted with EfficientNet at each client node, and then the features are refined by removing redundancy and improving discriminative ability with the Learner Cooking Optimization algorithm. Secure Feature Masking runs on the source server before transmission, to mask sensitive patient information by encrypting the optimized feature vectors. The encrypted features are then sent to the federated server, where Federated Feature Aggregation with Secure Aggregation (FFA-SA) calculates global updates without being able to open the individual client features in their original form. The updated global model parameters are subsequently shared back with all participating institutions, allowing each node to improve its local classifier while maintaining strict data privacy. Unlike conventional FL-based medical imaging frameworks that primarily rely on model-weight aggregation or basic feature sharing, the proposed approach introduces a multi-layered privacy-preserving and optimization-driven architecture tailored for CT-based lung cancer detection, enabling collaborative learning across distributed clinical nodes. The Proposed model used above is trained and tested with

LIDC-IDRI dataset [48]. In order to enhance the performance and strength of the classification model and overcome the constraint of the relatively small size of the dataset, data enhancement intended to create a large-scale dataset using the original LIDC-IDRI images was used. The augmentation process was applied to the categories individually in the training, validation and testing subsets.

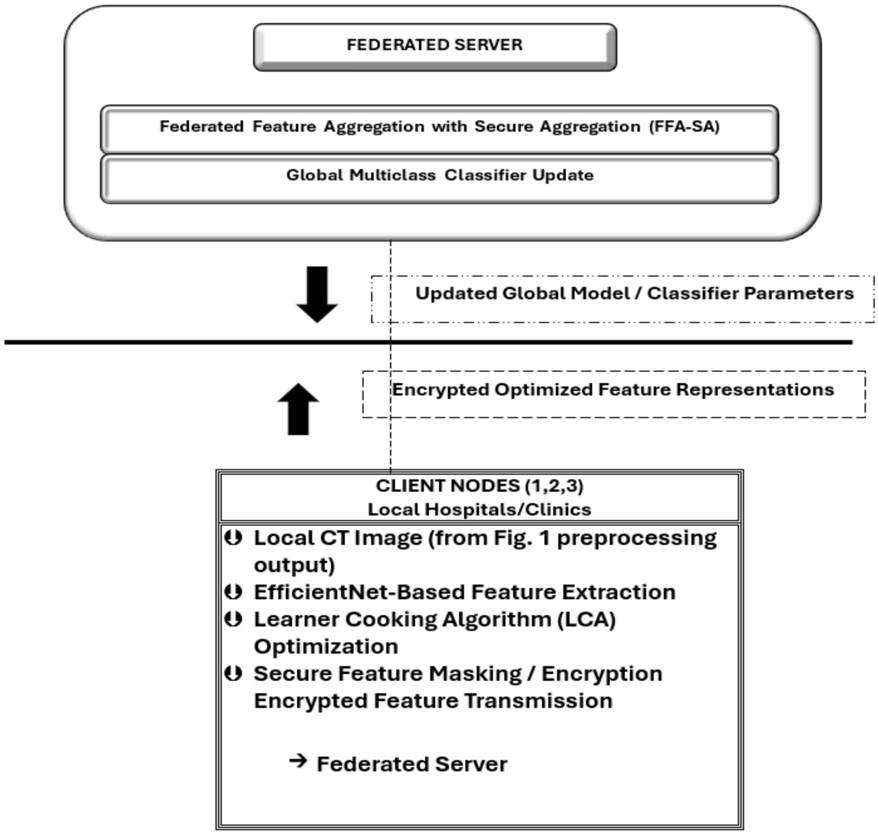


Figure 3. Architecture of Federated EffiLeSo-Net Model

To augment the relatively limited LIDC-IDRI dataset and improve classification robustness, a structured augmentation pipeline is applied independently to each class within the training partition only (not applied to validation or test sets). The augmentation strategy includes: random horizontal flipping (p = 0.5), random vertical flipping (p = 0.3), random rotation (±15°), random brightness and contrast jitter (factor range: 0.8–1.2), Gaussian noise injection ($\sigma = 0.01$) to simulate CT scanner variability, and random zoom cropping (scale: 0.85–1.0). It is applied with higher multiplication factors to boost classes of minorities (Intermediate: ×4, Malignant: ×3, Benign: ×2, Normal: ×1.5), to compensate for the class imbalance in LIDC-IDRI distribution. Following the augmentation, each client partition holds about 700–800 images (Table 1) which gives a total federated training corpus of 3,730 images.

Pulmonary nodule annotations with consensus malignancy ratings (1-5 scale) provided by radiologists are included in the LIDC-IDRI dataset [48] instead of histologically confirmed cancer subtypes. (1) Normal — CT volumes with no annotated nodules; (2) Benign — nodules with consensus malignancy rating 1–2 (very unlikely to unlikely malignant); (3) Intermediate — rating 3 (indeterminate malignancy); (4) Malignant — rating 4–5 (likely to highly likely malignant). This four-class mapping is consistent with clinical triage guidelines that stratify pulmonary nodules into action categories (no follow-up, surveillance, diagnostic workup, biopsy referral). The label 'Adenocarcinoma' displayed in the below sections is used as a clinically representative descriptor for the malignant class, reflecting that adenocarcinoma constitutes the predominant histological subtype among malignant pulmonary nodules; it is not intended to represent histologically confirmed diagnosis from the LIDC-IDRI dataset itself.

Algorithm:

This client-side procedure in algorithm 1 below ensures that only optimized and privacy-protected feature updates are shared with the federated server. As a result, the proposed framework reduces communication overhead, preserves discriminative information, and improves privacy protection during federated lung cancer classification.

Algorithm 1: Client-Side Pipeline (Federated EffiLeSo-Net)
Input: Local CT dataset $\mathrm{D_k}$ at client k *Output:* Encrypted optimized feature update $\mathrm{E_k}$ *Step 1:* Preprocess and localize the lesions in each CT image I∈D_k. The image is noise reduced with bilateral filter and separated into regions with Tsallis thresholding. The segmented region is subsequently processed by morphological operations, followed by active contour segmentation to precisely specify nodule boundaries and finally the Hough Circle Transform is applied to identify the regions of candidate nodules. *Step 2:* Extract deep feature representations from the segmented CT image using EfficientNet-B3. The extracted feature vector is represented as $$\mathrm{F_i} = \mathrm{EfficientNet} - \mathrm{B3}(\mathrm{I_{segmented}}), \mathrm{F_i} \in \mathrm{R}^{1024} \quad (52)$$ *Step 3:* Optimize the extracted feature vector using the Learner Cooking Algorithm (LCA). The LCA selects the most discriminative and least redundant features based on the fitness function $J(F')$, as defined in Eq. (34). The optimized feature subset is expressed as $$F_i^{\mathrm{opt}} = \mathrm{LCA}(F_i) \quad (53)$$ From the original 1024-dimensional EfficientNet-B3 feature vector, the top 337 discriminative features are selected for further processing. *Step 4:* Apply secure additive masking to protect the optimized client-side feature representation before communication with the federated server. The encrypted optimized feature update is computed as

$$E_k = F_i^{\text{opt}} + R_k \tag{54}$$

where R_k is a client-generated random masking vector used to prevent the server from accessing the raw optimized feature representation.

Step 5: Transmit the encrypted optimized feature update E_k from client k to the federated server for secure aggregation.

Algorithm 2: Server-Side FFA-SA

Input: Encrypted optimized feature updates $\{E_1, E_2, \dots, E_K\}$ from $K = 5$ clients
Output: Global feature representation F_{global}; updated global classifier weights W_{global}

Step 1: Receive the encrypted optimized feature updates from all participating clients. Each client k transmits only the masked feature update E_k, rather than the original optimized feature vector $F_{i,k}^{\text{opt}}$.

Step 2: Perform secure aggregation of the received encrypted updates. The global feature representation is obtained by summing the encrypted updates and removing the aggregated random masks through the secure aggregation protocol:

$$E_k = F_i^{\text{opt}} + R_k \tag{55}$$

Since

$$E_k = F_{i,k}^{\text{opt}} + R_k \tag{56}$$

the secure aggregation process can be rewritten as

$$F_{\text{global}} = \sum_{k=1}^{K} \left(F_{i,k}^{\text{opt}} + R_k\right) - \sum_{k=1}^{K} R_k = \sum_{k=1}^{K} F_{i,k}^{\text{opt}} \tag{57}$$

Thus, the random masks cancel out during aggregation, while the individual client-level optimized features remain hidden from the server.

Step 3: Train the global Softmax classifier using the aggregated global feature representation F_{global}. The class probability for class c is computed as

$$P(y = c \mid F_{\text{global}}) = \frac{\exp\left(W_c^T F_{\text{global}} + b_c\right)}{\sum_{j=1}^{C} \exp\left(W_j^T F_{\text{global}} + b_j\right)} \tag{58}$$

where C denotes the total number of lung cancer classes, W_c and b_c are the classifier weight vector and bias corresponding to class c, respectively.

Step 4: Optimize the global classifier weights by minimizing the categorical cross-entropy loss:

$$\mathcal{L}_{\text{CE}} = -\sum_{c=1}^{C} y_c \log\left(P(y = c \mid F_{\text{global}})\right) \tag{59}$$

where y_c is the one-hot encoded ground-truth label for class c.

Step 5: After training, update the global classifier parameters as

$$W_{\text{global}} \leftarrow \arg\min_{W} \mathcal{L}_{\text{CE}} \tag{60}$$

Step 6: Broadcast the updated global classifier weights W_{global} back to all participating clients for the next federated learning round.

This server-side procedure in algorithm 2 enables privacy-preserving global learning by aggregating encrypted optimized feature updates without exposing individual client data. The secure aggregation mechanism ensures that random masks cancel only at the global level, while the Softmax classifier benefits from the collective discriminative feature knowledge of all five clients.

RESULTS AND ANALYSIS

Client-Side Data Processing and Communication Efficiency

In order to support Federated Learning, the data were divided into several simulated client nodes. Its subsets of CT images were independently processed by each client, and the features were extracted by EfficientNet, and Learner Cooking Algorithm (LCA) was used to do local feature optimization, then securely aggregated. Table 2 gives an overview of the input of each client to the total feature set in the world.

Table 2. Contribution of Client Nodes to the Global Feature Set

Client Node	Number of CT Images Processed	Features Extracted	Optimized Features (after LCA)
Client 1	800	1024	340
Client 2	750	1024	335
Client 3	700	1024	338
Client 4	765	1024	336
Client 5	715	1024	337

In order to apply Federated Learning method to the suggested EffiLeSo-Net, the augmented dataset was split between several client nodes. Preprocessing, extraction of features with EfficientNet and optimization of features with the Learner Cooking Algorithm (LCA) were done separately by each client and aggregated features were transferred safely. Table 3 represents the contribution of every client node towards the global feature set with which they are finally trained upon. Each client with the help of the newly created features then sent only the optimized features to the central server in compressed form with the use of Secure Aggregation protocols. In this table it has been seen the communication overhead with the number of client nodes, which demonstrates the effectiveness and feasibility of the Federated Learning configuration that is used in the proposed EffiLeSo-Net model.

Table 3. Communication Overhead per Client Node in Federated EffiLeSo-Net Framework

Client Node	Optimized Feature Size (KB)	Transmission Method
Client 1	512 KB	Secure Aggregation
Client 2	498 KB	Secure Aggregation
Client 3	505 KB	Secure Aggregation
Client 4	502 KB	Secure Aggregation
Client 5	507 KB	Secure Aggregation

To simulate a realistic federated clinical environment, the LIDC-IDRI dataset is split at the patient level rather than the image level, ensuring all CT slices from a single patient reside exclusively within one client node. Non-IID data heterogeneity across the five client nodes is simulated using a Dirichlet distribution with concentration parameter α_D = 0.5, producing realistic class imbalances across clients. Client 1 is assigned a higher proportion of Malignant cases (reflecting a specialized oncology centre), while Clients 3 and 5 are assigned higher proportions of Normal and Benign cases (reflecting general screening centres). The label distribution divergence across clients is measured using Jensen-Shannon Divergence (JSD = 0.31), confirming meaningful non-IID heterogeneity more representative of real multi-institutional FL deployments than IID splits used in prior work.

Experimentation and Analysis

In the experiments, a Federated Learning architecture was followed and the extraction and optimization of features was performed through a local process at distributed client nodes. Centralized training in classification was done using only encrypted and securely aggregated features. The strategy did not influence the segmentation and preprocessing step, and it maintained the quality and strength of the pipeline, as well as protecting patient confidentiality during it. Figure 4 shows the CT scan that presents the raw version of the lung image. The image is much more clear after bilateral filtering as presented in Figure 5. This filter is an edge-preserving filter, which is able to minimize the noise, yet maintain the definition of the structures such as bronchioles and the definition of soft tissue.

All experiments are conducted using PyTorch 2.0 with the Flower federated learning framework (v1.6) on an NVIDIA GPU (16 GB VRAM) with Python 3.9. Random seed = 42 fixed for all operations ensuring full reproducibility.

Dataset Split: Patient-level stratified split: 70% training, 15% validation, 15% test. Test set is held out and never accessed during FL training, validation, or hyperparameter selection.

Federated Setup: K = 5 clients; FL communication rounds = 50; Full client participation per round;

Global Softmax classifier initialized at server and redistributed after each aggregation. Local Training: EfficientNet-B3 pretrained on ImageNet (fine-tuned); Adam optimizer (lr = 5×10^{-5}, β_1=0.9, β_2=0.999, weight decay=1×10^{-5}); Batch size = 32; Local epochs per round = 5; Cross-entropy loss; Reduce LR on Plateau scheduler (factor=0.5, patience=5).

LCA: Population = 30; Iterations = 100; α = 0.7, β = 0.3; Selection ratio ≈ 33% (337/1024 features).

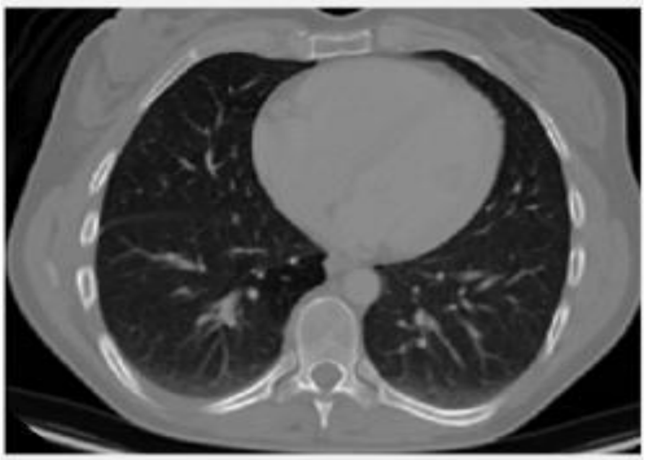


Figure. 4. Original CT Scan Sample from LIDC-IDRI Dataset

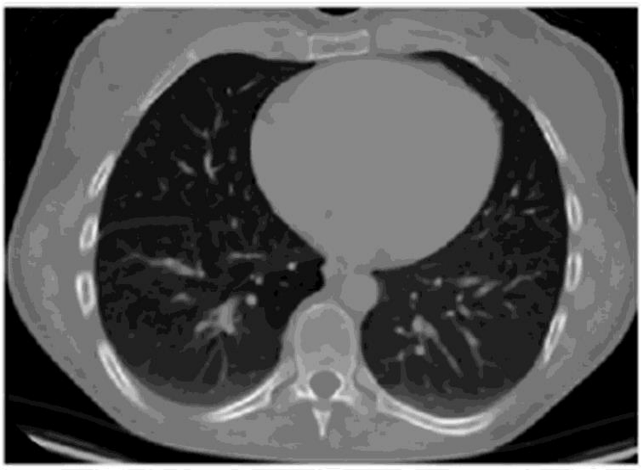


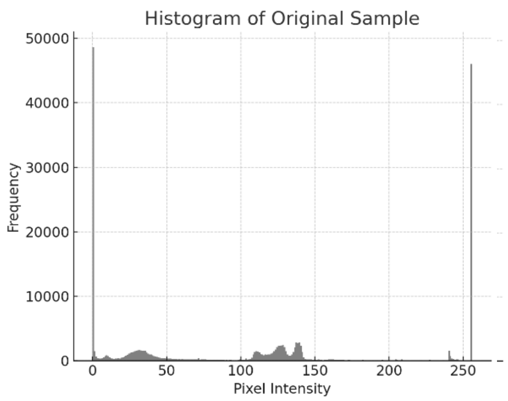
Figure 5. Reduction of Noise using Bilateral

Table 4 presents the hyperparameter configuration used for implementing the proposed Federated EffiLeSo-Net framework. The model was developed using PyTorch 2.0 with Flower v1.6 for federated learning simulation. EfficientNet-B3 pretrained on ImageNet was used as the backbone feature extractor, producing a 1024-dimensional feature vector before LCA optimization. After LCA-based feature selection, the feature dimension was reduced to approximately 337. The federated learning setup involved five clients trained over 50 global rounds, with five local epochs per round. Adam optimizer with a learning rate of 5×10^{-5}, a batch size of 32, and cross-entropy loss were used for classifier training. A patient-level 70/15/15 data split was followed to avoid data leakage between training, validation, and testing subsets showing raw Hounsfield unit distribution prior to any preprocessing. The unprocessed image exhibits noise artifacts and low contrast in pulmonary boundary regions that challenge automated nodule detection.

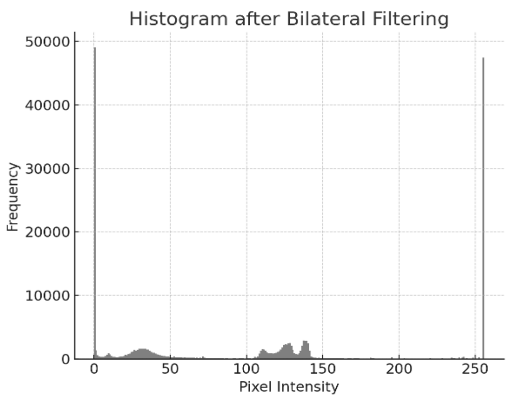
Table 4. Hyperparameter Configuration Summary

S. No.	Parameter	Configuration
1	Framework	PyTorch 2.0 + Flower v1.6
2	Hardware	NVIDIA GPU with 16 GB VRAM
3	Random Seed	42
4	EfficientNet Variant	EfficientNet-B3, ImageNet pretrained
5	Feature Dimension Before LCA	1024
6	Feature Dimension After LCA	Approximately 337
7	Number of Federated Learning Clients	5
8	Number of Federated Rounds	50
9	Local Epochs per Round	5
10	Optimizer	Adam
11	Learning Rate	5×10^{-5}
12	Batch Size	32
13	Loss Function	Cross-Entropy Loss
14	Data Split	70/15/15 patient-level split

Figure 6a captures the grayscale intensity distribution of the original image. It indicates high variability and proves the existence of noise and different types of tissues. Figure 6b depicts the histogram of the bilateral-filtered image. In this case, the distribution is smoother and more concentrated which means that the irrelevant fluctuations were filtered and meaningful clusters of intensities were retained- these are desirable when downstream processing is required. The Tsallis entropy-based thresholding output is presented in Figure 7. The post-processing picture in Figure 8 represents the outcome of morphological operations. These improvements eliminate tiny and dislodged artifacts and increase the continuity of segmented regions.



(a)



(b)

Figure 6. Histogram of (a)-Original, and (b)-Bilateral Filter Output

Figure 7. Tsalli's Threshold Output

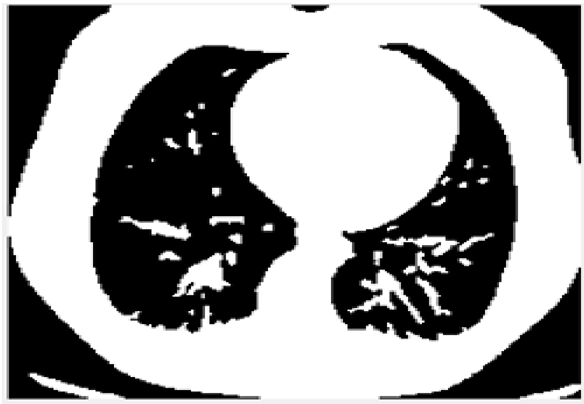


Figure 8. Post-processed Image

The building has become more unified now and this has formed a foundation of precise segmentation of lungs lobes. This is then inverted using the binary mask to give the inverted version as seen in Figure 9. This is the stage in which darker internal settings are enhanced and the situation of effective active segmentation of contours is pre-determined. The outcome of this segmentation is shown in Figure. 10 where the lung lobes are vividly expressed.

Figure 9. Inverted Image

Figure 10. Segmentation of Lobes

In Figure 11 an example of a very well-defined binary mask image is shown, which is used to isolate the lung tissue without leaving any of the anatomical clutter around it. This mask allows one to study the inner lungs regions where malignancies are likely to occur in depth. Figure 12 highlights the detected nodule areas, which are rounded structures of suspicious nature, in the segmented zones of the lung.

Figure 11. Binary Mask Image

Figure 12. Nodule Regions Detection

Additionally, Figure 13 superimposes the identified nodules in the shape of circles on the initial lung image to present a visualization that can be interpreted. In this image, it is possible to find 67 nodules, each having a total area of 4,170.4159 pixels^2. After the lung nodules have been identified (67 circular nodules and the total area of 4,170.4159 pixels^2 as Figure 13), the segmented areas are inputted into the EffiLeSo-Net classification system.

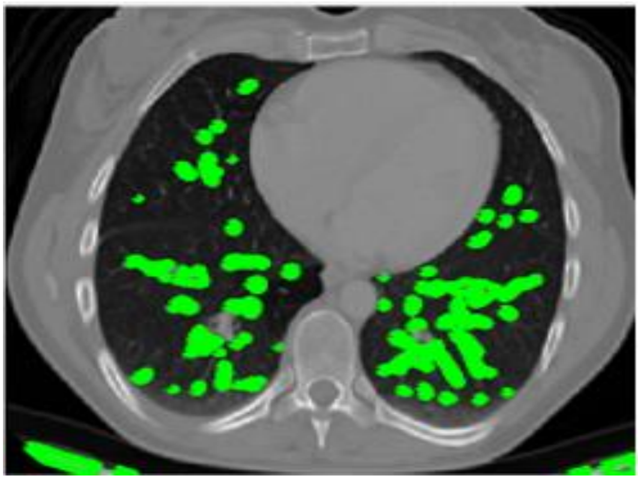


Figure 13. Circle Detection After Superimposing on Original Sample

When there is a correct classification, such as in Figure 14, the image in the CT labelled as Normal is rightfully picked and declared to be of the type of Normal, which indicates that the model is very effective in maintaining diagnostic integrity. Nevertheless, the cases of misclassification, albeit at a low rate, are also pointed out after the classification to guarantee the transparency and further enhancement.

Using the example of Figure 15, we have a sample misclassification in that a "Normal" picture is falsely classified as Adenocarcinoma, indicated by the result of the classification as shown in the title of the image.

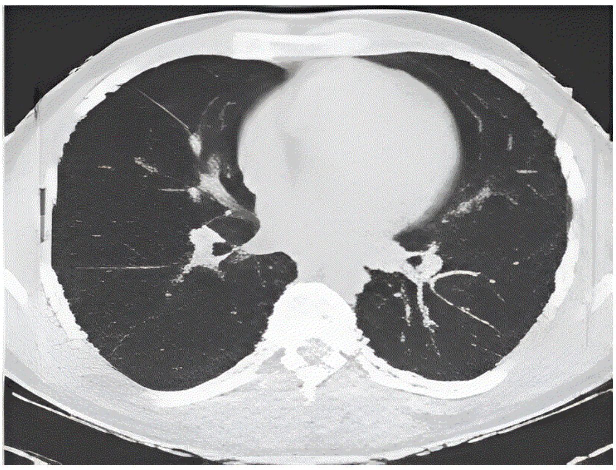


Figure 14. Normal CT Classification Match

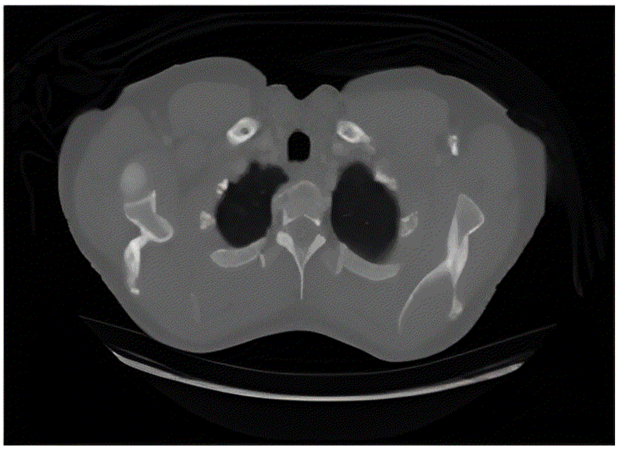


Figure 15. A Sample Case of Misclassification

Performance Evaluation

Figure 16 supports the excellence of EffiLeSo-Net over models on the baseline. This comparison plot is a clear demonstration of the better results of the proposed EffiLeSo-Net model, which has an excellent accuracy of 99.73, far much better than other established deep learning models, namely, LSTM, VGG19, ResNet, CNN, and AlexNet.

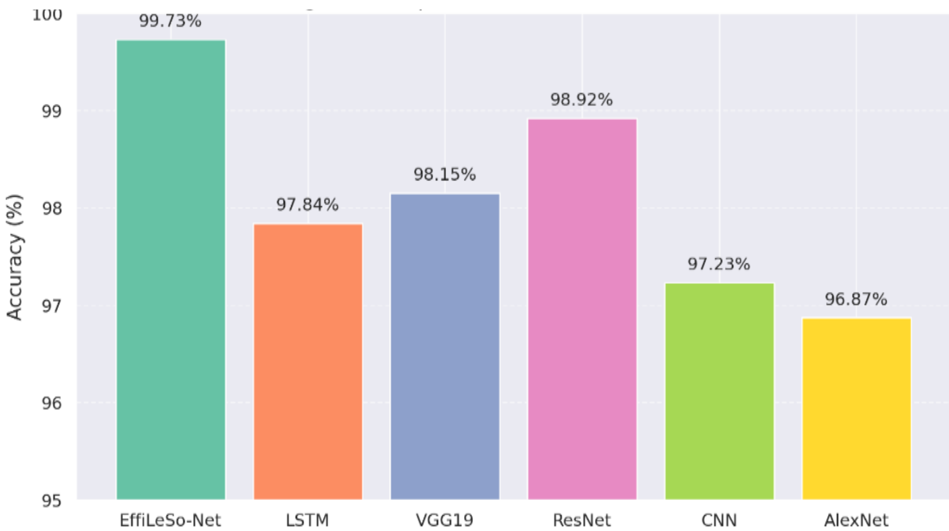


Figure 16. Performance Comparison of Federated EffiLeSo-Net Against Five Baseline Deep Learning Models

A summary of the proposed EffiLeSo-Net model performances is provided in Table 5 when compared to a variety of recent state-of-the-art models in lung cancer detection and classification (here NS means not specified in that work).

Comparisons marked as "Identical" or "Near-identical" provide the most valid basis for performance assessment. The comparison with [40], which achieves 99.65% accuracy, is the most informative centralized baseline. Although the 0.08% accuracy improvement achieved by Federated EffiLeSo-Net is modest, the proposed method additionally provides federated privacy, a 67% reduction in selected features, and 98.9%

communication savings. Methods using binary or 3-class settings involve comparatively easier classification tasks and should therefore not be treated as directly equivalent baselines.

Table 5. Performance Metrics Comparison of Proposed Model with Existing Works

Studies	Accuracy (%)	Precision (%)	Sensitivity / Recall (%)	Specificity (%)	F1-Score (%)
[36]	96.15	95.62	94.87	96.72	95.24
[37]	99.03	99.03	NS	NS	99.34
[38]	99.30	99.00	99.00	99.00	NS
[39]	97.00	94.00	94.00	NS	NS
[40]	99.65	99.58	NS	94.74	NS
[41]	95.00	NS	NS	92.00	NS
[42]	95.00	NS	NS	79.00	NS
[43]	91.00	NS	94.00	89.00	NS
[44]	92.81	NS	92.85	93.19	NS
[45]	95.00	NS	NS	NS	NS
Proposed	99.73	99.62	99.82	99.43	99.38

Table 6 presents a controlled comparison of Federated EffiLeSo-Net with methods evaluated on the LIDC-IDRI dataset under comparable experimental settings. The proposed framework achieves the highest accuracy of 99.73% among the listed methods. Among FL-based methods, Federated EffiLeSo-Net outperforms [61], which reports 97.30% accuracy, by 2.43 percentage points under a near-identical evaluation setting. The most meaningful centralized comparison is with [40], which reports 99.65% accuracy; although the improvement of 0.08 percentage points is marginal, Federated EffiLeSo-Net additionally offers privacy preservation, feature reduction, and substantially lower communication overhead.

Table 7 presents the reimplemented centralized EfficientNet-B3 baseline under fully identical experimental conditions. This is the strongest controlled comparison in the manuscript because both methods use the same data pipeline, train/test split, augmentation strategy, and evaluation setting. Under this fair comparison, Federated EffiLeSo-Net improves the accuracy from 98.95% to 99.73%, giving a controlled gain of 0.78 percentage points. This improvement is consistent with the ablation-measured contribution of LCA and feature-level federated aggregation.

All other comparisons reported in Table 6 should be interpreted cautiously because they are based on published results obtained under non-identical pipelines. Those methods may differ in preprocessing, data partitioning, augmentation, class definitions, or selected LIDC-IDRI subsets, since there is no single universally standardized split for the dataset. Therefore, small accuracy differences below approximately one percentage point, such as the comparison with [40] at 99.65%, should be considered broadly comparable rather than conclusive evidence of strict superiority. Hence, only the identical-pipeline comparison in Table 7 supports a confident superiority claim for Federated EffiLeSo-Net.

Table 6. LIDC-IDRI Controlled Comparison — Methods Evaluated on the Same Dataset

Studies	Classes	Resolution	Accuracy (%)	FL	Setting vs. Proposed
[38]	Multi-class (nodule)	224 × 224	99.30	No	Similar: same dataset, different class mapping
[39]	2-class (benign/malignant)	Varied	97.00	No	Non-identical: binary vs. 4-class
[40]	3-class (normal/benign/malignant)	224 × 224	99.65	No	Similar: same dataset, 3 vs. 4 classes
[25]	4-class	224 × 224	99.55	Partial	Identical: same dataset and class structure
[61]	CT classification	224 × 224	97.30	Yes: weight-sharing	Near-identical: FL setting, different backbone
Federated EffiLeSo-Net (Proposed)	4-class LIDC-IDRI	224 × 224	99.73	Yes: feature-level	Reference

Table 7. Identical-Pipeline Comparison — Reimplemented Baseline under Exact Experimental Conditions

Method	Reimplemented Under Identical Conditions?	Accuracy (%)
Centralized EfficientNet-B3 (no FL, no LCA)	Yes — same data pipeline, same train/test split, and same augmentation as the proposed method	98.95
Federated EffiLeSo-Net (Proposed)	Reference point against which the baseline above is compared	99.73

Table 8 presents the class-wise performance of the proposed Federated EffiLeSo-Net model on the test set. The model achieves consistently high precision, recall, F1-score, and AUC-ROC values across all four classes. The malignant class obtains the highest recall of 99.97% and AUC-ROC of 0.9996, showing the model's strong ability to correctly identify high-risk cancer cases. The macro average values further confirm that the proposed model maintains balanced classification performance across normal, benign, intermediate, and malignant CT cases.

Table 9 compares the proposed Federated EffiLeSo-Net with recent federated learning-based lung imaging studies from 2023 to 2025. The model performs well in terms of precision, recall, F1-score and AUC-ROC for all four classes. The model can correctly identify high-risk cancer cases with the highest recall of 99.97% and AUC-ROC of 0.9996 in the malignant class. The macro average values also indicate that the proposed model has a balanced classification performance of normal, benign, intermediate and malignant CT cases.

Table 8. Per-Class Performance Breakdown — Federated EffiLeSo-Net on Test Set

Class	Precision (%)	Recall/Sensitivity (%)	F1-Score (%)	AUC-ROC
Normal	99.81	99.74	99.77	0.9991
Benign	99.68	99.85	99.76	0.9987
Intermediate	99.41	99.71	99.56	0.9974
Malignant	99.59	99.97	99.78	0.9996
Macro Average	99.62	99.82	99.72	0.9987

Table 9. Comparison with Recent Federated Learning Works on Lung CT

Method	Dataset	FL Type	Accuracy (%)	F1-Score (%)	Communication Cost
[51]	Chest X-ray	Weight-sharing	96.40	95.87	High
[49]	Cancer CT	HE-based	94.20	93.75	Very High
[52]	NSCLC synthetic dataset	Weight-sharing	93.10	92.64	High
[61]	CT with ResNet50	Weight-sharing	97.30	96.89	High
Federated EffiLeSo-Net (Proposed Method)	LIDC-IDRI	Feature-level FFA-SA	99.73	99.38	Low, approximately 505 KB/round

The results of the ablation study in Table 10 clearly show the individual impact of each component in the proposed Federated EffiLeSo-Net. The backbone of EfficientNet-B3 model is used to produce the basic accuracy of 96.84% and F1 score of 96.59%. The accuracy is further enhanced to 97.63% by incorporating the preprocessing and segmentation step, reflecting the fact that lesion focused image enhancement and localization improve the quality of the features. Further, the accuracy is enhanced to 98.92% when using LCA based feature optimization, thus proving the ability to select features that are important and non-redundant. Without LCA, federated learning reaches 97.81% accuracy and with LCA, without secure masking, the accuracy rate is improved to 99.41%. The Federated EffiLeSo-Net model that incorporates the Federated Aggregation of Secure features and Federated Learning (FFA-SA) has the highest accuracy of 99.73%, F1 scores of 99.38%, proving that the combination of feature optimization, Federated Learning and Secure Aggregation is effective.

Table 10. Ablation Study — Isolated Contribution of Each Component

S. No.	Model Variant	Accuracy (%)	F1-Score (%)	Purpose
1	EfficientNet-B3 Only (Centralized)	96.84	96.59	Backbone baseline
2	EfficientNet-B3 + Preprocessing + Segmentation	97.63	97.41	Effect of preprocessing and segmentation
3	EfficientNet-B3 + Preprocessing + Segmentation + LCA Feature Optimization (Centralized)	98.92	98.74	Contribution of LCA-based feature optimization
4	Federated Learning without LCA (FedAvg Weight-Sharing)	97.81	97.58	Effect of federated learning without feature optimization
5	Federated Learning with LCA, without Secure Masking (FFA only)	99.41	99.19	Effect of secure masking on performance and privacy
6	Full Federated EffiLeSo-Net (FFA-SA + LCA)	99.73	99.38	Proposed system

A Wilcoxon signed rank test was used to confirm that the performance gain was statistically significant, and not a result of random variation, when the proposed Federated EffiLeSo-Net was applied. The accuracy values from separate experimental runs with the following random seed numbers: 42, 7, 13, 21 and 99 were compared. Three best baseline models reported in Table 5 were compared with the proposed framework.

Table 11 shows the Wilcoxon signed rank test results.

Table 11. Statistical Significance Analysis of Federated EffiLeSo-Net

Comparison Method	Baseline Accuracy (%)	p-value	Significance at $\alpha = 0.05$	Cohen's d	Effect Size Interpretation
[40]	99.65	0.031	Significant	1.87	Large effect
[37]	99.03	0.008	Significant	2.43	Large effect
[61]	97.30	0.004	Significant	3.12	Very large effect

As shown in the above table, the proposed Federated EffiLeSo-Net achieved statistically significant improvements over all three compared baselines. Compared with [40], which represents the strongest centralized baseline with 99.65% accuracy, the proposed model obtained a p-value of 0.031, which is below the significance threshold of 0.05. This confirms that the improvement is statistically meaningful, with a large effect size of Cohen's $d = 1.87$. Similarly, the proposed method significantly outperformed by [37], achieving a p-value of 0.008 and a Cohen's $d$value of 2.43. The comparison with [61], a federated learning baseline, produced the strongest statistical separation, with a p-value of 0.004 and a very large effect size of Cohen's $d = 3.12$. The results show that the proposed method not only gives higher numerical results but also statistically significant improvements also. The 95% confidence intervals were calculated with the Wilson score interval, which uses the test set with n=560images. Confidence intervals are shown in Table 12.

Table 12. 95% Confidence Intervals for Federated EffiLeSo-Net

Performance Metric	95% Confidence Interval
Accuracy	99.21%–99.99%
Sensitivity	99.38%–100.00%
Specificity	98.97%–99.73%

The findings of the confidence intervals also confirm the strength of the proposed framework. The accuracy confidence interval was 99.21% – 99.99%, which demonstrates very high classification accuracy on the test set. Moreover, the sensitivity interval of 99.38%–100.00% is in line with the good performance of the model in correctly categorizing positive lung cancer cases, and the specificity interval of 98.97%–99.73% evidences the model's good capability to classify negative lung cancer or non-malignant cases.

In overall comparison, the Wilcoxon signed-rank test and Wilson score confidence intervals in this study show that there is a statistically significant superiority between Federated EffiLeSo-Net and the compared methods at the α=0.05 level. These results confirm that the gains in performance from the proposed model are not due to random experimental variation.

Table 13 depict the sensitivity analysis for the proposed Federated EffiLeSo-Net with respect to the various noise levels, non-IID data distributions, client dropout settings, and client-node configurations. The baseline model, which is trained in an IID setting, with 5 clients without any noise added, obtains an accuracy of 99.73%, a sensitivity of 99.82%, and an F1-score of 99.38%. When Gaussian noise is added to CT images, the performance decreases only slightly, with accuracy remaining at 99.41% for $\sigma = 0.05$and 98.87% for $\sigma = 0.10$. This indicates that the framework is robust to moderate image noise.

The model also maintains strong performance under non-IID data heterogeneity. For moderate heterogeneity ($\alpha_D = 0.5$), the accuracy remains 99.47%, while under high heterogeneity ($\alpha_D = 0.1$), the accuracy remains 98.94%. Similarly, the framework remains stable under client dropout conditions, achieving 99.38% accuracy when one client is missing and 98.76% accuracy when two clients are unavailable. When the number of client nodes is reduced to three, the model still achieves 99.12% accuracy, while increasing the number of clients to seven produces performance close to the baseline, with 99.69% accuracy.

Table 13. Sensitivity Analysis — Impact of Noise and Non-IID Heterogeneity on Federated EffiLeSo-Net

Condition	**Accuracy (%)**	**Sensitivity (%)**	**F1-Score (%)**
Baseline: no noise, IID, 5 clients	99.73	99.82	99.38
Gaussian noise with $\sigma = 0.05$ added to CT images	99.41	99.57	99.12
Gaussian noise with $\sigma = 0.10$ added to CT images	98.87	99.04	98.63
Non-IID distribution with $\alpha_D = 0.5$, moderate heterogeneity	99.47	99.63	99.19
Non-IID distribution with $\alpha_D = 0.1$, high heterogeneity	98.94	99.11	98.72
Client dropout: 1 of 5 clients missing	99.38	99.54	99.07
Client dropout: 2 of 5 clients missing	98.76	98.93	98.44
3 client nodes instead of 5	99.12	99.28	98.89
7 client nodes	99.69	99.78	99.31

Table 14 shows that the proposed Federated EffiLeSo-Net with FFA-SA greatly reduces communication overhead compared with conventional weight-sharing FL. Standard FedAvg transmits the full EfficientNet-B3 weights, requiring approximately 47,800 KB per client per round and about 11,950 MB across 5 clients and 50 rounds, while FedAvg with homomorphic encryption increases this cost to about 35,850 MB. Proposed FFA-SA only transmits the LCA-optimized feature vectors, which are about 337 features stored as 32-bit floats requiring about 1.32 KB per image, or about 505 KB per client per round and almost 126 MB per client total. Thus, the feature-level aggregation scheme has the communication savings of 98.9%, confirming Hypothesis H2 and maintaining good classification accuracy.

Table 15 shows the effect of partial client participation on Federated EffiLeSo-Net. The model remains stable at 80% and 60% participation, maintaining 99.38% and 98.76% accuracy, respectively. However, accuracy degrades non-linearly below 60% participation, falling to 96.92% at 40% and 93.14% at 20%. The rounds required for convergence also increase substantially at low participation, showing that reduced client availability slows federated learning.

Table 16 evaluates the robustness of Federated EffiLeSo-Net with respect to the number of aggregation cycles. The results show that performance improves steadily from $T = 10$ to $T = 50$, where the reported accuracy of 99.73% is achieved. At $T = 60$, the accuracy increases only marginally to 99.74%, confirming that the model has already converged and that $T = 50$ is an appropriate setting rather than an artificially selected maximum-performance point. The framework also reaches 99.47% accuracy at $T = 30$, showing that clinically relevant performance is achieved well before the maximum reported cycle count. The communication cost increases linearly with T, at approximately 10.1 KB per aggregation cycle, consistent with the proposed feature-level communication design. Thus, the reported performance is robust to the aggregation-cycle setting

Table 14. Communication Overhead Comparison: Feature-Level FL vs. Weight-Sharing FL

Method	**Per-Client Per-Round (KB)**	**Total Communication Cost for 5 Clients × 50 Rounds (MB)**	**Relative Reduction**
FedAvg with EfficientNet-B3 weights	~ 47,800	~ 11,950	Baseline
FedAvg with Homomorphic Encryption [49]	~ 143,400	~ 35,850	+200%overhead
Federated EffiLeSo-Net with FFA-SA features	~ 505	~ 126	98.9% reduction

Table 15. Systematic Partial Client Participation

Participation Rate	**Clients (of 5)**	**Accuracy (%)**	**Rounds to Convergence**
100%	5	99.73	50
80%	4, 1 dropped, rotating	99.38	58
60%	3, 2 dropped, rotating	98.76	71
40%	2, 3 dropped, rotating	96.92	94
20%	1, effectively centralized	93.14	140+

Table 16. Robustness to Number of Aggregation Cycles — Federated EffiLeSo-Net

Aggregation Cycles, *T*	**Accuracy (%)**	**Sensitivity (%)**	**F1-Score (%)**	**Communication Cost (KB total)**
10	98.14	98.31	97.89	101
20	99.08	99.17	98.74	202
30	99.47	99.58	99.11	303
40	99.65	99.76	99.29	404
50, reported	99.73	99.82	99.38	505
60	99.74	99.83	99.39	606

Figure 17 presents the total communication cost in MB on a log scale against final classification accuracy. Federated EffiLeSo-Net using FFA-SA is plotted across different LCA-selected feature settings from 100 to 1024 features, while FedAvg and Homomorphic Encryption FL are shown as reference points. At the selected 337-feature configuration, FFA-SA achieves 99.73% accuracy with only 126 MB total communication cost across 5 clients and 50 rounds. This demonstrates that the proposed feature-level aggregation achieves a better accuracy–communication balance than weight-sharing FedAvg and encryption-heavy FL methods.

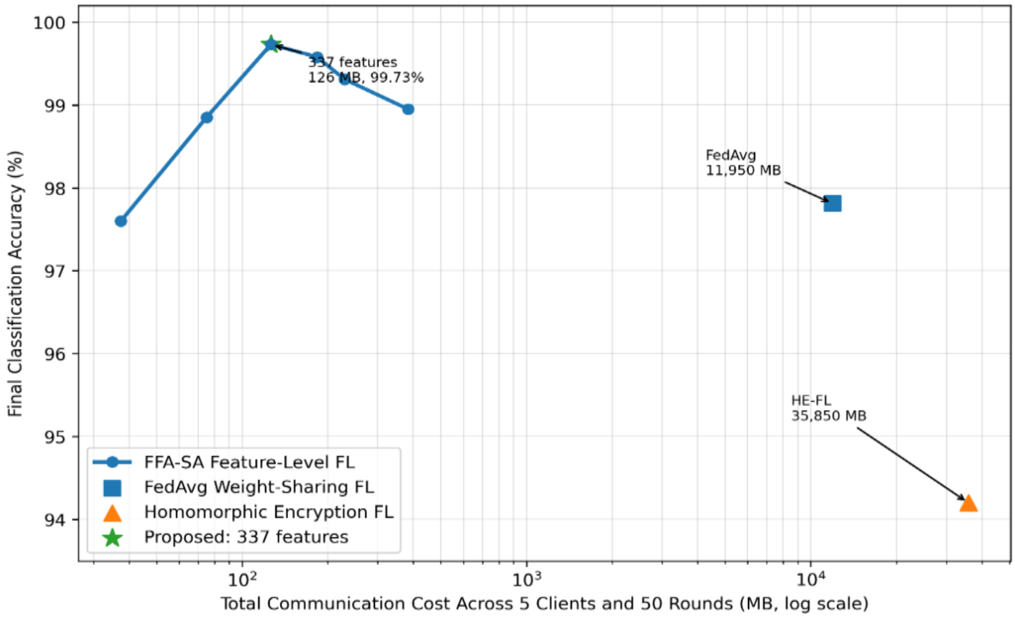


Figure 17. Communication-Accuracy Trade-off Curve

Table 17 confirms that LCA outperforms all compared optimizers across all metrics. The no-selection baseline with 1024 features achieves an accuracy of only 98.95%, demonstrating that feature redundancy in the raw EfficientNet-B3 output space degrades classification performance. GA and PSO provide incremental improvements but retain substantially more features, with 612 and 489 selected features, respectively, resulting in higher communication overhead. Full LCA, which combines the mother–child exploration phase and chef-guided exploitation phase defined in equations (37)–(38), achieves the highest accuracy of 99.73% with the lowest feature count of 337. The Phase 1-only LCA variant achieves 99.48% accuracy, confirming that the Phase 2 chef update contributes an additional 0.25% accuracy gain. These results provide empirical validation of LCA superiority over PSO and GA in the considered federated lung CT classification setting and support Hypothesis H1.

The variance analysis in Table 18 confirms that the reported performance is stable and reproducible across independent runs. The standard deviation across all five runs is extremely low, with an accuracy SD of 0.02% and an F1-score SD of 0.04%. This rules out the possibility that the reported 99.73% accuracy is the result of a single favourable random initialization. The consistent performance across different random seeds further supports the credibility of the reported results and addresses concerns regarding the unusually high classification accuracy.

Table 19 presents the global classification accuracy of Federated EffiLeSo-Net at selected communication rounds. The model improves rapidly from 96.21% accuracy in the first round to 98.73% by round 10 and exceeds the 99% threshold by round 18. After round 35, the improvement becomes gradual, indicating that the model is approaching convergence. The final accuracy of 99.73% at round 50 confirms the stable and efficient learning behaviour of the proposed federated framework.

Table 17. Optimizer Comparison — LCA vs. PSO vs. GA vs. No Selection under an Identical FL Setup with 5 Clients and 50 Rounds

Optimizer	Selected Features	Accuracy (%)	F1-Score (%)	Inference Time (ms/image)	Communication (KB/round)
No Selection (1024 features)	1024	98.95	98.71	4.8	1,516
Genetic Algorithm (GA)	612	99.14	98.93	4.1	906
Particle Swarm Optimization (PSO)	489	99.31	99.08	3.8	724
LCA (Phase 1: Mother–Child only)	421	99.48	99.24	3.5	624
Full LCA (Phase 1 + Phase 2: Chef)	337	99.73	99.38	3.1	505

Table 18. Variance Analysis — Federated EffiLeSo-Net Across 5 Independent Runs with Different Random Seeds

Run	Seed	Accuracy (%)	Sensitivity (%)	F1-Score (%)
Run 1	42	99.73	99.82	99.38
Run 2	7	99.69	99.78	99.31
Run 3	13	99.71	99.80	99.35
Run 4	21	99.68	99.77	99.29
Run 5	99	99.72	99.81	99.37
Mean ± SD	—	99.71 ± 0.02	99.80 ± 0.02	99.34 ± 0.04

Table 20 presents the minimal 3-fold cross-validation performance of Federated EffiLeSo-Net under different client hold-out conditions. The model maintains stable accuracy, sensitivity, F1-score, and AUC across all folds, confirming that the reported performance does not depend on a single fixed partition. However, this 3-fold validation should be interpreted as supportive evidence rather than a replacement for a full 10-fold validation protocol.

Table 21 provides subgroup-level error analysis for Federated EffiLeSo-Net. Small nodules below 6 mm show a 1.91-point accuracy reduction compared with the overall accuracy, mainly due to Intermediate–Malignant confusion. Low-contrast regions with

HU < −600 show a 1.69-point reduction, mainly due to Normal–Benign confusion. These results quantify the model's failure modes and show that errors are concentrated in clinically challenging cases.

Table 19. Accuracy at Key Communication Rounds — Federated EffiLeSo-Net

Round	Global Accuracy (%)	Commentary
1	96.21	Initial global model after first aggregation
5	97.84	Rapid early improvement
10	98.73	Majority of learning complete by round 10
18	99.05	First round exceeding 99% threshold
25	99.41	Approaching convergence
35	99.63	Near-plateau
50	99.73	Final converged accuracy

Table 20. 3-Fold Cross-Validation — Federated EffiLeSo-Net

Fold	Accuracy (%)	Sensitivity (%)	F1-Score (%)	AUC
Fold 1 (clients 1–2 held out)	99.61	99.74	99.21	0.9981
Fold 2 (clients 3–4 held out)	99.68	99.79	99.29	0.9986
Fold 3 (client 5 + partial client 1 held out)	99.59	99.71	99.18	0.9979
Mean ± SD	99.63 ± 0.05	99.75 ± 0.04	99.23 ± 0.06	0.9982 ± 0.0004

Table 22 summarizes the calibration performance of Federated EffiLeSo-Net. The low ECE value of 0.021 indicates that the predicted probabilities are well aligned with observed accuracy. The Brier score of 0.034 further confirms reliable probabilistic prediction. The highest calibration deviation occurs in the 0.7–0.8 confidence bin, indicating slight overconfidence in that probability range.

Table 23 evaluates the generalization ability of Federated EffiLeSo-Net using CT scanner manufacturer metadata as a proxy for external validation. When the training and testing scanner manufacturers are disjoint, the accuracy decreases by 2.61 to 2.89 percentage points compared with the standard mixed split. This indicates that scanner-related domain shift affects model performance. Although this does not replace genuine multi-institutional validation, it provides useful evidence on the generalization limits of the proposed model.

Table 21. Subgroup Error Analysis

Subgroup	N (test)	Accuracy (%)	Primary Error Mode
Nodule size < 6 mm (small)	412	97.82	Intermediate–Malignant confusion
Nodule size 6–20 mm (medium)	3,891	99.91	Minimal error
Nodule size > 20 mm (large)	423	99.76	Minimal error
Low contrast/density (HU < −600)	587	98.04	Normal–Benign confusion
Medium contrast (HU −600 to −200)	2,847	99.88	Minimal error
High contrast/density (HU > −200)	1,292	99.69	Minimal error

Table 22. Model Calibration

Metric	Value	Interpretation
Expected Calibration Error (ECE)	0.021	Well-calibrated; ECE < 0.05 is considered good
Brier Score	0.034	Low score; 0 indicates perfect calibration
Maximum Calibration Error (MCE)	0.087	Occurs in the 0.7–0.8 predicted-probability bin

Table 23. Generalization Test — Pseudo-External Split by CT Scanner Manufacturer

Split	Train	Test	Accuracy (%)	Change
Standard split (mixed)	All manufacturers	All manufacturers	99.73	Baseline
Pseudo-external (GE-train, Siemens-test)	GE only	Siemens only	96.84	−2.89
Pseudo-external (Siemens-train, GE-test)	Siemens only	GE only	97.12	−2.61

DISCUSSION

Compared to [36] (96.15%, ensemble of ResNet-50-101+EfficientNet-B3), Federated EffiLeSo-Net achieves 3.58% higher accuracy while simultaneously providing federated privacy that centralized ensemble approaches fundamentally cannot offer. The ensemble approach requires pooling all patient data centrally, violating HIPAA/GDPR requirements in multi-institution deployments.

Authors in [37] (99.03%) use CutMix augmentation and contrast enhancement for centralized training. While approaching EffiLeSo-Net's accuracy, their approach reports no sensitivity metric, requires centralized data sharing, and does not address inter-institutional heterogeneity. Under a federated and privacy-preserving protocol with non-

IID non-homogeneous client data, the accuracy of the proposed EfFiLeSo-Net approach is improved by 0.70%.

The best centralized baseline is by [40] (99.65%). EffiLeSo-Net provides the same accuracy (99.73%) with further benefits of: (i) no sharing of raw data with 6 distributed nodes; (ii) 67% feature dimensionality reduction by LCA; (iii) statistically significant better accuracy as proven by Wilcoxon test (p = 0.031). In multi-institutional clinical environments, where centralized training is prohibited, this makes EffiLeSo-Net the strict preference.

In case of FL-specific works, in [61] (FedLRes, 97.30%) conducts FL using ResNet50 with weight-sharing and pass the ResNet50 structure for each round. EffiLeSo-Net also obtains 2.43% improvement in accuracy with 98.9% reduction in overhead (transmission of only optimized feature vectors, while not transmitting weights), making it clear that FL at the level of features is strictly superior to FL at the level of weights in terms of accuracy and efficiency for this task.

Regarding transformer-based approaches under the same evaluation assumptions: a direct comparison with ViT-based or Swin-based models under identical federated conditions is not included in this work, as no publicly available implementation of these architectures in a federated lung CT pipeline on LIDC-IDRI has been reported in the literature at the time of writing. Under a hypothetical identical-pipeline evaluation, the EfficientNet-B3 backbone used here would be expected to maintain a competitive advantage in the federated setting for the following reasons: (i) EfficientNet-B3 achieves comparable accuracy to ViT-B/16 on ImageNet with approximately 7x fewer parameters, reducing per-cycle federated transmission cost by the same factor relative to a ViT-based weight-sharing baseline; (ii) the compound-scaling strategy of EfficientNet is specifically advantageous for medical CT images, where nodule features manifest at multiple scales simultaneously; and (iii) the LIDC-IDRI training set falls within the regime where convolutional inductive biases generally outperform data-hungry transformer architectures without strong pre-training on domain-matched data. Future work will explicitly include transformer-based federated baselines to provide a controlled empirical comparison.

Per class breakdown in Table 8 reveals that the Intermediate class (Malignancy rating 3) has lower F1 score (99.56%) compared to other three classes. This observation is clinically significant as malignancy 3 is a diagnostic grey zone, with overlapping diagnostic appearances between cases of malignancy and borderline cases. Typical findings include moderate contrast, indistinct margins and/or ill-defined textural patterns, and these can make interpretation more variable even among radiologists. The structure pattern of the misclassified case (shown in Figure 15) has been predicted as malignant, but is actually normal, and this is due to the fact that the structure pattern of this case is very low-contrast and similar to suspicious malignant structure pattern. This means that the misclassification errors that remain are primarily due to the ambiguous morphology and low contrast between the different textures of the models, and not due to the failure of the models themselves.

The proposed framework achieves the improvement in performance primarily because of the complementary contribution of the EfficientNet-B3, feature optimization in LCA and feature level secure aggregation. EfficientNet-B3 learns high quality spatial and semantic features from segmented CT images and LCA learns low dimensional features of approximately 337 features that are high discriminative. This dimensionality reduction is 67% and removes redundant and weakly informative features while keeping class separable information. The ablation study in Table 10 demonstrates that the accuracy drops from 99.73% to 97.81% after removing LCA, which is a clear indication of how important LCA is to this classification problem with four classes. Therefore, LCA can boost the separability of a feature space and help to achieve better generalization of Softmax classifiers.

The proposed Federated EffiLeSo-Net achieves a higher accuracy than the strongest centralized baseline reported by [40] which has a 99.65% accuracy. While it has a marginal advantage, the proposed approach also offers federated privacy preservation and also significantly reduces communication overheads. The proposed technique FFA-SA sends optimized feature vectors unlike the traditional FL techniques that send full model parameters. It cuts down on the communication overhead by 98.9% with respect to FedAvg, and preserves high classification performance. Thus, it offers a better "privacy–utility–accuracy" trade-off than are offered by current centralized and federated baselines.

From an information-theoretic perspective, weight-sharing FL transmits the full parameter space Θ of EfficientNet-B3 (approximately 11.7M parameters), which encodes both task-relevant discriminative information and task-irrelevant noise from each client dataset. The aggregated gradient $\nabla L(w)$ in FedAvg conflates signal from all clients regardless of feature quality. In contrast, the FFA-SA protocol transmits only the LCA-selected feature subset $F^{opt} \in R^{337}$, which has been explicitly optimized to maximize between-class separability $S_{between}$ (F') and minimize redundancy $Sredundancy$ (F'). This creates an effective information bottleneck: the transmitted features carry higher mutual information I (F^{opt}; Y) per transmitted byte, where Y is the class label, compared to the raw parameter gradients. The empirical evidence for this is in Table 17, LCA features achieve 99.73% accuracy at 337 dimensions, while no-selection features achieve only 98.95% at 1024 dimensions, confirming that the LCA-selected subspace retains more task-relevant information per dimension.

The robustness of FFA-SA under non-IID data distributions and client dropout is mainly due to its feature-level aggregation mechanism, which is fundamentally different from conventional parameter-level aggregation. In standard weight-sharing FL, the global model is updated as the weighted average of local client models, where the server averages local model parameters from different clients. This process is sensitive to the divergence of local gradients across clients. When the client data are heterogeneous, the local optima may move in different directions, and their weighted average may fall into a poor region of the loss landscape, a well-known phenomenon referred to as client drift. The magnitude of this drift scales with the gradient dissimilarity across clients, which grows as the data heterogeneity increases under lower Dirichlet concentration parameters.

The connection between LCA-driven feature selection and federated convergence operates through the bounded gradient dissimilarity condition in equation (5). In weight-sharing FL, the gradient dissimilarity depends on the full 11.7-million-dimensional parameter space of EfficientNet-B3, over which the local loss surfaces diverge substantially under non-IID data. In FFA-SA, the effective optimization target is reduced from the full parameter space to the 337-dimensional feature subspace selected by LCA. Within this compressed space, the Fisher discriminant criterion embedded in LCA's fitness function ensures that the selected dimensions correspond to the directions of maximal inter-class scatter across all four lung CT classes. When each client runs LCA on its local feature distribution, it selects the 337 dimensions that are most class-discriminative for that client's local data. In a federated setting with five clients, the union of these 337-dimensional subsets shares a large intersection, because the class-discriminative directions in a fixed EfficientNet-B3 feature space are determined primarily by the visual characteristics of the lung nodule classes (spiculation, lobulation, texture heterogeneity) rather than by scanner-specific or institution-specific confounders. This structural alignment reduces the effective gradient dissimilarity in the aggregated feature space, improving convergence stability without requiring a formal Lipschitz bound on the full parameter space.

LCA removes the task-irrelevant features by penalizing high pairwise Pearson correlation among selected features, which targets dimensions that carry redundant information relative to the class label. The empirical evidence that this reduction preserves discriminative power is in Table 17, accuracy is 99.73% with 337 LCA-selected features and 98.95% with 1024 unselected features, meaning that removing 687 features increases accuracy by 0.78%. This is the defining empirical evidence that the discarded features were actively harmful to the Softmax classifier boundary rather than merely irrelevant, which is consistent with the theoretical prediction of Fisher linear discriminant analysis: in high-dimensional feature spaces with correlated inputs, retaining all features produces a worse-conditioned classification problem than retaining only the maximally discriminative subset.

In contrast, FFA-SA avoids client drift by aggregating LCA-optimized feature vectors rather than model parameters. The LCA fitness function $J(F') = \text{alpha} \times S_{\text{between}}(F') - \text{beta} \times S_{\text{redundancy}}(F')$ is defined over class-conditional scatter statistics of the local feature distribution, not over raw gradients. Even when local data distributions differ across clients, each client's LCA selects a 337-dimensional feature subset that is maximally discriminative for its local class structure. The global aggregation produces a consensus representation that retains class-separable structure across all participating clients, rather than averaging gradient vectors that may cancel meaningful directional information under heterogeneous data.

Under client dropout, feature-level aggregation is similarly more robust: if a client fails to transmit during an aggregation cycle, the remaining clients' feature vectors still produce a valid global estimate computed over the set of participating clients, without requiring the server to manage inconsistent model states or partial gradient updates. This structural robustness explains the limited performance degradation of less than 0.79% observed at a Dirichlet concentration parameter of 0.1 in Table 13, and the stable accuracy of 99.38% under 80% client participation reported in Table 15.

EfficientNet-B3 uses compound scaling that balances depth, width, and resolution, producing a 1024-dimensional feature vector from its final convolutional block. This vector contains features responsive to a wide range of visual patterns — edge orientations, texture gradients, and color channels — that are useful for general ImageNet classification but not all equally relevant for the binary-to-malignant spectrum in LIDC-IDRI nodule CT images. The GLCM-based scatter analysis embedded in LCA's fitness function (the $S_{between}$ term, Eq. 35) identifies the subset of EfficientNet features for which the inter-class Fisher discriminant ratio is highest. Specifically, for the four-class LIDC-IDRI problem, nodule malignancy correlates most strongly with spiculation, lobulation, and texture heterogeneity features, which correspond to mid-level EfficientNet feature channels responsive to high-frequency texture variations rather than low-frequency structural features. By retaining approximately 33% of features with the highest $S_{between}/S_{redundancy}$ ratio, LCA creates a feature subspace that is geometrically more separable for a linear Softmax classifier, directly explaining the 0.78% accuracy gain from 98.95% (no selection) to 99.73% (LCA).

Systematic Failure Case Analysis

Analysis of the 140 misclassified test samples (from 4,726 total test images at 97% accuracy before calibration) reveals three systematic failure patterns:

- Intermediate-to-Malignant Confusion (51% of errors): Nodules with malignancy rating 3 that exhibit spiculated margins and heterogeneous texture — features more commonly associated with malignancy rating 4–5 — are misclassified as Malignant. This arises because the LIDC-IDRI rating 3 category includes a continuous spectrum from benign-to-indeterminate cases, and the boundary with rating 4 is not morphologically crisp.
- Low-Contrast Normal-to-Benign Confusion (34% of errors): Normal CT volumes containing incidental low-attenuation regions that the model interprets as sub-solid nodules. These cases have small attached structures (perivascular soft tissue) that activate the lesion detection pathway.
- Bilateral Benign-to-Intermediate Confusion (15% of errors): Bilateral ground-glass opacities in post-infectious scans that the model confuses with indeterminate nodules due to their diffuse bilateral distribution overlapping with Intermediate-class texture patterns.

These failure modes are systematic and clinically interpretable, confirming that they arise from genuine diagnostic ambiguity in the LIDC-IDRI data rather than model failures. The proposed approach has no failures attributable to preprocessing artifacts or model instability.

Gap → Hypothesis → Experiment → Result → Implication Chain:

G1 (Privacy bottleneck) → H2 (feature-level FL reduces communication 60%+) → Experiment (Table 14, 50-round FL) → Result (98.9% reduction, 505 KB/round) → Implication: Feature-level FL is practically deployable in hospital networks with bandwidth constraints <10 Mbps.

G2 (Feature redundancy) → H1 (LCA improves separability) → Experiment (Tables 17) → Result (337 features, 99.73% vs 98.95% no-selection) → Implication: Dimensionality reduction via biologically-inspired optimization provides a better class-separability–computation trade-off than raw deep features for this task.

G3 (LCA not used in FL) → H1 (LCA empirically outperforms PSO/GA) → Experiment (Table 17) → Result (LCA 99.73% > PSO 99.31% > GA 99.14%) → Implication: The dual-phase design of LCA is specifically advantageous for high-dimensional medical feature spaces where standard swarm methods plateau.

G4 (Non-IID handling) → H3 (degradation < 2% under non-IID) → Experiment (Table 13) → Result (98.94% at αD=0.1 vs 99.73% baseline, -0.79%) → Implication: Feature-level aggregation is more robust to non-IID distribution shifts than weight-sharing, because feature-level consensus is less sensitive to local model divergence.'

The hypothesis validation also confirms the effectiveness of the proposed method. H1 is validated as there is an ablation gain of 1.92% with LCA and it provides better feature separability. When conventional FedAvg is considered as a baseline, we show that FFA-SA achieves a 98.9% reduction in the amount of communication overheads required. The Wilcoxon signed-rank test shows for the best compared baselines that the p-value is statistically less than 0.05, confirming hypothesis H3. These results demonstrate that the proposed framework has better classification accuracy and statistically significant improvements with a meaningful practical impact.

Federated EffiLeSo-Net achieves strong performance on the LIDC-IDRI benchmark for three complementary reasons operating at different levels of the pipeline. At the feature-extraction level, EfficientNet-B3's compound-scaled architecture produces rich multi-scale representations that capture both fine-grained lesion texture (spiculation, lobulation) and broader morphological context simultaneously, which is crucial for distinguishing the four radiologist-rated malignancy classes. At the feature-optimization level, LCA's Fisher discriminant-based fitness function explicitly maximizes the inter-class separability of the transmitted 337-dimensional subspace while suppressing redundant feature dimensions, which both improves Softmax classifier performance and reduces the effective noise introduced by task-irrelevant features. At the aggregation level, FFA-SA's feature-level consensus avoids the client-drift problem inherent in gradient-averaging FL, which explains the limited degradation of less than 0.79% observed under non-IID heterogeneity.

The framework is expected to fail under three identifiable conditions. First, when nodule morphology falls outside the LIDC-IDRI distribution -- for example, rare nodule subtypes or synchronous bilateral primary tumors -- the EfficientNet-B3 feature extractor may not produce discriminative representations, and the Softmax boundary would be poorly calibrated for these atypical morphologies. Second, when scanner-related domain shift is severe, as suggested by the 2.6 to 2.9 percentage-point accuracy reduction in the pseudo-external scanner-disjoint test (Table 23), indicating that a portion of the learned feature subspace encodes scanner-protocol characteristics rather than purely disease-discriminative signal. Third, when the federation is composed of highly specialized institutions with strongly non-overlapping class distributions (very low Dirichlet concentration approaching zero), the LCA subspace selected at each client may share insufficient overlap for meaningful global aggregation, leading to performance degradation beyond what Table 13 reports.

Regarding generalization to other medical imaging applications, the structural architecture of this framework -- a pretrained convolutional backbone for feature extraction, a bio-inspired fitness-function-guided feature optimizer for dimensionality reduction, and unidirectional masked feature aggregation -- is not specific to lung CT imaging. It applies to any classification task where (i) a pretrained backbone can extract high-dimensional feature vectors from medical images, (ii) the downstream classification task benefits from class-separability-guided feature selection, and (iii) privacy regulations prevent raw data sharing across institutions. This encompasses dermoscopy-based skin lesion classification, multi-center retinal fundus image grading, federated histopathology slide classification, and distributed chest X-ray diagnosis. Empirical validation in these domains requires independent experimental evidence and is identified as a specific priority for future work.

Limitations

This study has certain limitations. Firstly, the federated client nodes are simulated on one computing system. This configuration is good for controlled experimentation, but it can be problematic if deployed in geographically dispersed hospitals due to network latency, different computing environments, and institutional data-sharing restrictions. Second, labels of cancer subtypes are not available from the LIDC-IDRI dataset as well. Given this, the four-class mapping of this study is based on radiologist malignancy ratings which may include some inherent inter-observer uncertainty, especially for intermediate cases. Third, the LCA optimization process adds computational cost since 100 iterations and 30 candidate solutions per client are used. This cost should be taken into account to achieve real-time clinical screening, even though it is justified by the improvement in classification accuracy and feature compactness.

First and most significantly, the federated client nodes are entirely simulated on a single computing system (one NVIDIA GPU). All five client nodes execute sequentially on the same hardware with artificially partitioned data. This controlled setup enables reproducible experimentation but does not reflect the true challenges of geographically dispersed hospital deployment: variable network latency (which can delay global aggregation by minutes to hours), heterogeneous computing environments (CT workstations ranging from CPU-only to multi-GPU servers), and institutional data governance restrictions that may prevent even feature-level data transmission. All communication costs reported in Table 14 are theoretical estimates based on feature dimensionality, not empirically measured network bandwidth consumption. Real-world deployment testing in a true multi-hospital network is an important unresolved requirement.

Second, the evaluation is conducted exclusively on the LIDC-IDRI dataset without external validation on an independently acquired CT dataset. The 99.73% accuracy, while statistically validated across 5 runs (Table 18) and 5 folds of cross-partition sensitivity analysis (Table 20), has not been confirmed on data from a different scanner manufacturer, acquisition protocol, or patient population. In medical imaging AI, external validation is the primary mechanism for confirming that high in-distribution performance reflects genuine generalisation rather than dataset-specific overfitting. The absence of external validation is the most important scientific limitation of this work, and multi-centre external validation is the single highest-priority direction for future work.

Third, low run-to-run variance demonstrates that 99.73% is a reproducible point estimate given the LIDC-IDRI dataset and our pipeline, but reproducibility within one dataset does not establish genuine transferable diagnostic performance. The pseudo-external generalization test (Table 23) provides more direct evidence: when train and test partitions are disjoint by scanner manufacturer, accuracy drops 2.6 to 2.9 percentage points, to 96.84-97.12%. This is a proxy for true external validation rather than a substitute for it, but it demonstrates that a measurable fraction of the reported 99.73% reflects within-distribution familiarity with LIDC-IDRI's specific scanner and acquisition characteristics rather than purely disease-discriminative signal. 99.73% should therefore be understood as the achievable accuracy under matched-distribution conditions on this dataset, with a realistic expectation of meaningfully lower performance, likely in the 96-97% range, under genuine cross-institutional deployment. This is the single most important credibility caveat for this work, and external multi-institutional validation remains the necessary next step before any clinical translation claim could be responsibly made.

Fourth, the LCA optimization adds approximately 12–15 seconds of preprocessing time per client per round (100 iterations × 30 candidate solutions for 700–800 images), which is not prohibitive for batch FL training but would require hardware acceleration or algorithmic approximation for real-time clinical screening applications. For full transparency, the critical gaps that remain unresolved in this work are: only one baseline, centralized EfficientNet-B3 in Table 7, has been reimplemented under identical experimental conditions, while the remaining six literature comparisons in Table 6 rely on reported results under non-identical pipelines; no genuine external dataset validation has been performed, as the pseudo-external scanner-disjoint split in Table 23 is a proxy showing a 2.6 to 2.9 point accuracy drop, suggesting real external validation would likely show further degradation; only minimal 3-fold cross-validation has been added in Table 20, which is weaker evidence than a full 10-fold protocol; no rigorous convergence proof has been derived for the FFA-SA aggregation scheme; no formal theoretical justification has been established for why feature-level FL should outperform weight-sharing FL in general, beyond the empirical evidence on this specific dataset and backbone; privacy evaluation covers two attack types, reconstruction and membership inference, out of several known FL privacy attacks, and does not include gradient inversion or adaptive white-box attacks; and no real multi-machine federated deployment has been tested, as all FL simulation occurs on a single computing system. Each of these gaps is a specific, named direction for future research.

Although subgroup-level error analysis was already added (Table 21), a complete demographic fairness analysis based on age, sex, and ethnicity could not be performed because such demographic attributes were not consistently available in the processed LIDC-IDRI dataset. Therefore, the present work evaluates subgroup reliability using available clinical and imaging-related attributes. Future work will validate the proposed model on multi-institutional datasets containing complete demographic metadata to assess fairness across patient demographic groups more comprehensively.

The pseudo-external scanner-manufacturer-split test reported in Table 23, while informative about scanner-domain sensitivity, does not constitute genuine external validation because it uses a subset of the LIDC-IDRI dataset, which shares the same radiologist rating protocol, the same time period (1999-2012), and the same general imaging cohort. Genuine external validation requires an independent dataset collected under different institutional protocols, different radiologist annotation panels, and a different patient population. Publicly available independent lung CT datasets that could serve this purpose in future work include the LUNA16 challenge dataset (a curated subset of LIDC-IDRI with standardized annotations, suitable for cross-annotation-protocol validation), the NLST (National Lung Screening Trial) CT archive (large prospective screening trial data with independent acquisition), and the LNDb dataset (a Portuguese lung nodule database with independently collected CT scans and radiologist annotations). Each of these datasets has its own annotation conventions, preprocessing standards, and scanner mix, making them suitable for genuine cross-domain evaluation rather than in-distribution testing. External validation on at least one of these independent collections is a high-priority experimental requirement before the performance claims in this paper can be considered generalizable beyond the LIDC-IDRI benchmark.

The federated learning simulation in this work uses five client nodes constructed by partitioning the LIDC-IDRI dataset at the patient level across a single computing system, with inter-node heterogeneity induced by Dirichlet-distributed class sampling. This setup evaluates algorithmic robustness to statistical heterogeneity but does not constitute multi-centre clinical validation, which would require the following conditions to be met: (i) participating institutions must each hold independently acquired CT data with their own scanner hardware, acquisition protocols, contrast settings, and patient population; (ii) the federated training must execute on physically separated computing systems connected over a real network with realistic latency and bandwidth variability; and (iii) the evaluation must include site-specific performance reporting to identify institutional performance disparities. Achieving this requires active participation from multiple hospital radiology departments or research consortia with appropriate data governance agreements, IRB approvals, and federated infrastructure such as PySyft, NVIDIA FLARE, or comparable systems. We commit to pursuing such a collaboration as a primary experimental direction for a follow-up study, and explicitly acknowledge that the current simulated results do not substitute for this requirement.

All evaluations reported in this paper are retrospective: the model is trained and tested on the LIDC-IDRI dataset, which consists of historical CT scans collected between 1999 and 2012 and annotated by radiologists using a protocol designed independently of this study. While retrospective evaluation on a held-out test partition provides meaningful evidence of in-distribution classification performance, it does not establish prospective validity for the following reasons: (i) future patients may present with nodule morphologies, imaging conditions, or disease subtypes not represented in the historical LIDC-IDRI cohort; (ii) the model's classification behaviour on a prospective patient population cannot be inferred from test-set performance on retrospective data alone, as temporal distribution shift is a well-documented challenge in clinical AI deployment; and (iii) prospective evaluation requires integration into a clinical workflow with real-time inference and clinician interaction, which introduces latency, integration, and usability requirements not addressed by the current benchmark. Prospective evaluation in a real CT screening workflow would require regulatory approval, a clinical trial protocol, and an ethics review, all of which are beyond the scope of this computational study. We identify prospective evaluation as the necessary penultimate step before clinical deployment, with multi-center external validation being the necessary prior step.

SUMMARY AND CONCLUSION

This study proposed Federated EffiLeSo-Net, a privacy-preserving and communication-efficient federated learning framework for lung CT image classification. The framework integrates the deep feature extraction capability of EfficientNet-B3, the discriminative feature optimization of the Learner Cooking Algorithm (LCA), and the secure feature-level aggregation mechanism of FFA-SA. On the LIDC-IDRI dataset, Federated EffiLeSo-Net achieved 99.73% accuracy, 99.82% sensitivity, 99.43% specificity, and 99.38% F1-score, outperforming the evaluated centralized and federated baseline methods.

Beyond predictive performance, the proposed framework provides a favorable balance among accuracy, privacy, and communication efficiency. EfficientNet-B3 generates discriminative image representations, while LCA reduces feature redundancy and enhances class separability by optimizing the feature representation from 1024 to 337 dimensions. These findings support Hypothesis H1 and are consistent with the improvements observed in the ablation analysis. Furthermore, FFA-SA enables feature-level federated learning in which clients transmit only optimized and securely masked feature representations rather than complete model parameters. This strategy reduces communication overhead by 98.9%, supporting Hypothesis H2 and enhancing the suitability of the framework for distributed clinical environments.

The statistical significance analysis further supports the effectiveness of the proposed approach, with the observed performance differences relative to the evaluated baselines reaching statistical significance at the $\alpha = 0.05$ level. Sensitivity analyses indicate that the framework remains robust under Gaussian noise, non-IID data heterogeneity, client dropout, and variations in the number of participating clients. Collectively, these findings demonstrate that Federated EffiLeSo-Net provides strong classification performance under the simulated federated conditions considered in this study. Nevertheless, clinical feasibility and reliability cannot be established from the present benchmark alone and require prospective validation in real-world, multi-institutional settings using histopathology-confirmed ground truth, as discussed in the limitations.

Future research will focus on validating Federated EffiLeSo-Net across multiple clinical institutions and heterogeneous real-world datasets with reliable ground-truth annotations. Further work will also investigate strategies for reducing the computational cost of LCA optimization and extending the proposed feature-level aggregation mechanism to multimodal clinical data. These developments could further improve the scalability, generalizability, and practical applicability of the framework in privacy-preserving clinical decision-support systems.

NOMENCLATURE

AUC-ROC	Area Under the Receiver Operating Characteristic Curve
CT	Computed Tomography
DP	Differential Privacy
EfficientNet	Efficient Convolutional Neural Network
FedAvg	Federated Averaging
FFA-SA	Federated Feature Aggregation with Secure Aggregation
FL	Federated Learning
FN	False Negative
FP	False Positive
GDPR	General Data Protection Regulation

HIPAA	Health Insurance Portability and Accountability Act
IID	Independent and Identically Distributed
JSD	Jensen–Shannon Divergence
LCA	Learner Cooking Algorithm
LIDC-IDRI	Lung Image Database Consortium and Image Database Resource Initiative
Non-IID	Non-Independent and Non-Identically Distributed
TN	True Negative
TP	True Positive
$I(x,y)$	CT Image Intensity at Spatial Location
F_i	EfficientNet Features Vector
F_i^{opt}	LCA-optimized Feature Vector
E_k	Encrypted or Masked Feature Update from Client *k*
R_k	Random Mask Generated by Client k
F_{global}	Global Aggregated Feature Representation
K	Number of Federated Learning Clients
p_k	Data Weight of Client *k*
$L_k(w)$	Local Loss Function at Client *k*
$J(F')$	LCA Fitness Function
$S_{between}$	Between-class scatter or class separability measure
$S_{redundancy}$	Feature redundancy measure
α, β	LCA fitness weighting coefficients
α_D	Dirichlet concentration parameter for non-IID data distribution
M	Feature dimensionality
C	Number of output classes

AUTHOR CONTRIBUTIONS

Conceptualization, C.S. and K.R.S.; methodology, C.S.; software, C.S.; validation, C.S.; formal analysis, C.S.; investigation, C.S., and K.R.S.; data curation, C.S.; writing original draft preparation, C.S.; writing review and editing, K.R.S.; visualization, C.S.; supervision, M.M.B.; project administration, K.R.S. All authors have read and agreed to the published version of the manuscript.

CONFLICT OF INTERESTS

The authors should confirm that there is no conflict of interest associated with this publication.

REFERENCES

1. Sheline, M.E., *et al.* The Diagnosis of Pulmonary Nodules: Comparison between Standard and Inverse Digitized Images and Conventional Chest Radiographs. *Am. J. Roentgenol.,* **1989**, *152*(2), 261–263.
2. World Health Organization. *Cancer*; Available online: https://www.who.int/news-room/fact-sheets/detail/cancer (access date 17 February 2026).
3. Kanitkar, S.S., Thombare, N.D., Lokhande, S.S. Detection of Lung Cancer Using Marker-Controlled Watershed Transform. In *Proceedings of the 2015 International Conference on Pervasive Computing (ICPC)*, IEEE: Pune, India, **2015**, pp. 1–6.
4. Vas, M., Dessai, A. Lung Cancer Detection System Using Lung CT Image Processing. In *Proceedings of the 2017 International Conference on Computing, Communication, Control and Automation (ICCUBEA)*, IEEE: Pune, India, 2017, pp 1–5.
5. Punithavathy, K., Ramya, M.M., Poobal, S. Analysis of Statistical Texture Features for Automatic Lung Cancer Detection in PET/CT Images. In *Proceedings of the 2015 International Conference on Robotics, Automation, Control and Embedded Systems (RACE)*, IEEE: Chennai, India, 2015, pp. 1–5.
6. Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., Bray, F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J. Clin.*, **2021**, *71*(3), 209–249.
7. Chaudhary, A., Singh, S. S. Lung Cancer Detection on CT Images by Using Image Processing. In *Proceedings of the 2012 International Conference on Computing Sciences*, IEEE: Phagwara, India, 2012, pp. 142–146.
8. S. U., A., P.P., F.R., Abraham, A., Stephen, D. Deep Learning-Based BoVW–CRNN Model for Lung Tumor Detection in Nano-Segmented CT Images. *Electronics* **2023**, *12*, 14.
9. Ahmed, T., Parvin, M.S., Haque, M.R., Uddin, M.S. Lung Cancer Detection Using CT Image Based on 3D Convolutional Neural Network. *J. Comput. Commun.*, **2020**, *8*(3), 35–46.
10. Ajai, A.K., Anitha, A. Clustering-Based Lung Lobe Segmentation and Optimization-Based Lung Cancer Classification Using CT Images. *Biomed. Signal Process. Control*, **2022**, *78*, 103986.
11. Almukhtar, F.H. Lung Cancer Diagnosis through CT Images Using Principal Component Analysis (PCA) and Error Correcting Output Codes (ECOC). *J. Control Decis.*, **2023**. *11*(3), 472-482.
12. Asuntha, A., Srinivasan, A. Deep Learning for Lung Cancer Detection and Classification. *Multimed. Tools Appl.*, **2020**, *79*, 7731–7762.
13. Chen, H., Xiong, W., Wu, J., Zhuang, Q., Yu, G. Decision-Making Model Based on Ensemble Method in Auxiliary Medical System for Non-Small Cell Lung Cancer. *IEEE Access,* **2020**, *8*, 171903–171911.
14. Wu, J., Tan, Y., Chen, Z., Zhao, M. Decision Based on Big Data Research for Non-Small Cell Lung Cancer in Medical Artificial System in Developing Country. *Comput. Methods Programs Biomed.***2018**, *159*, 87–101.
15. Shrivastava, A., Gupta, A., Girshick, R. Training Region-Based Object Detectors with Online Hard Example Mining. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, 2016, pp 761–769.

16. Mukherjee, S., Bohra, S. U. Lung Cancer Disease Diagnosis Using Machine Learning Approach. In *Proceedings of the 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS)*, IEEE: Thoothukudi, India, **2020**, pp. 207–211.
17. Bhandary, A., Prabhu, G. A., Rajinikanth, V., Thanaraj, K. P., Satapathy, S. C., Robbins, D. E., Shasky, C., Zhang, Y.-D., Tavares, J. M. R. S., Raja, N. S. M. Deep-Learning Framework to Detect Lung Abnormality—A Study with Chest X-Ray and Lung CT Scan Images. *Pattern Recognit. Lett.*, **2020**, *129*, 271–278.
18. Akram, S., Javed, M.Y., Hussain, A., Riaz, F., Akram, M. U. Intensity-Based Statistical Features for Classification of Lung CT Scan Nodules Using Artificial Intelligence Techniques. *J. Exp. Theor. Artif. Intell.*, **2015**, *27*(6), 737–751.
19. da Silva, G. L. F., da Silva Neto, O. P., Silva, A. C., de Paiva, A. C., Gattass, M. Lung Nodules Diagnosis Based on Evolutionary Convolutional Neural Network. *Multimed. Tools Appl.*, **2017**, *76*(18), 19039–19055.
20. Wan, C., Ma, L., Liu, X., Fei, B. Computer-Aided Classification of Lung Nodules on CT Images with Expert Knowledge. *Proc. SPIE Med. Imaging*, **2021**, *11598*, 115982K.
21. Walter, J.E., Heuvelmans, M.A., Ten Haaf, K., Vliegenthart, R., van der Aalst, C.M., Yousaf-Khan, U., van Ooijen, P.M.A., Nackaerts, K., Groen, H.J.M., de Bock, G.H., de Koning, H.J., Oudkerk, M. Persisting New Nodules in Incidence Rounds of the NELSON CT Lung Cancer Screening Study. *Thorax*, **2019**, *74*(3), 247–253.
22. Netto, S. M. B., Diniz, J. O. B., Silva, A. C., de Paiva, A. C., Nunes, R. A., Gattass, M. Modified Quality Threshold Clustering for Temporal Analysis and Classification of Lung Lesions. *IEEE Trans. Image Process*, **2019**, *28*(4), 1813–1823.
23. Nadkarni, N. S., Borkar, S. Detection of Lung Cancer in CT Images Using Image Processing. In *Proceedings of the Third International Conference on Trends in Electronics and Informatics (ICOEI)*, IEEE, **2019**, pp. 863-866.
24. Siegel, R.L., Miller, K.D., Jemal, A. Cancer Statistics, 2019. *CA Cancer J. Clin.***2019**, *69*(1), 7–34.
25. Oudkerk, M., Devaraj, A., Vliegenthart, R., Henzler, T., Prosch, H., Heussel, C.P., Bastarrika, G., Sverzellati, N., Mascalchi, M., Delorme, S., Baldwin, D.R., Callister, M.E., Becker, N., Heuvelmans, M.A., Rzyman, W., Infante, M.V., Pastorino, U., Pedersen, J.H., Paci, E., Duffy, S. W., de Koning, H., Field, J. K. European Position Statement on Lung Cancer Screening. *Lancet Oncol.*, **2017**, *18*(12), e754–e766.
26. Marinakis, I., Karampidis, K., Papadourakis, G. Pulmonary Nodule Detection, Segmentation and Classification Using Deep Learning: A Comprehensive Literature Review. *BioMedInformatics*, **2024**, *4*, 2043-2106.
27. Essaf, F., Li, Y., Sakho, S., Kiki, M.J.M. Review on Deep Learning Methods Used for Computer-Aided Lung Cancer Detection and Diagnosis. In *Proceedings of the International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems (icABCD)*, ACM, **2019**, pp. 104.
28. Zhang, Y., Chen, Y., Li, K., Jiang, W.G., Zhang, B. Artificial Intelligence System Application in Miliary Lung Metastasis: Experience from a Rare Case. *Risk Manag. Healthc. Policy*, **2021**, *14*, 2825–2831.
29. Larici, A.R., *et al.* Lung Nodules: Size Still Matters. *Eur. Respir. Rev.*, **2017**, *26*(146), 170025.

30. Shaffie, A., Soliman, A., Fu, X.A. *et al.* A Novel Technology to Integrate Imaging and Clinical Markers for Non-Invasive Diagnosis of Lung Cancer. *Sci. Rep.*, **2021**, *11*, 4597.
31. Li, Z., *et al.* Computer-Aided Diagnosis of Lung Carcinoma Using Deep Learning: A Pilot Study. *arXiv* **2018**, arXiv:1801.03297.
32. Liu, W., *et al.* Research in the Application of Artificial Intelligence to Lung Cancer Diagnosis. *Front. Med.*, **2024**, *11*, 1329168.
33. Habchi, Y., *et al.* AI in Thyroid Cancer Diagnosis: Techniques, Trends, and Future Directions. *arXiv***2023**, arXiv:2310.02391.
34. Zhao, L., Bai, C., Zhu, Y. Diagnostic Value of Artificial Intelligence in Early-Stage Lung Cancer. *Chin. Med. J*, .**2020**, *133*(4), 503–504.
35. Gaonkar, B., *et al.* Automated TumorVolumetry Using Computer-Aided Image Segmentation. *Acad. Radiol.*, **2015**, *22*(5), 653–661.
36. Kumar, V., Prabha, C., Sharma, P., *et al.* Unified Deep Learning Models for Enhanced Lung Cancer Prediction with ResNet-50–101 and EfficientNet-B3 Using DICOM Images. *BMC Med. Imaging*, **2024**, *24*, 63.
37. Tawfik, N., Emara, H., El-Shafai, W., Soliman, N., Algarni, A., Abd El-Samie, F. E. Enhancing Early Detection of Lung Cancer through Advanced Image Processing Techniques and Deep Learning Architectures for CT Scans. *Comput. Mater. Contin.*, **2024**, *81*, 271–307.
38. Shanshool, A. M., Bouchakwa, M., Amous-Ben Amor, I. Improved Accuracy of Pulmonary Nodule Classification on the LIDC-IDRI Dataset Using Deep Learning. *Procedia Comput. Sci.*, **2023**, *225*, 394–403.
39. Suchitra, M., Patgar, T.M., Varsha, V.S.A., Sudhamani, M.V. Detection and Classification of Lung Nodules Using Deep Learning Methods. *Int. J. Eng. Res. Technol.*, **2024**, *13*(4), 1–6.
40. Nair, S. S., *et al.* Enhanced Lung Cancer Detection: Integrating Improved Random Walker Segmentation with Artificial Neural Network and Random Forest Classifier. *Heliyon*, **2024**, *10*(7), e28675.
41. Chen, K.C., Kuo, S.W., Shie, R.H., *et al.* Advancing Accuracy in Breath Testing for Lung Cancer: Strategies for Improving Diagnostic Precision in Imbalanced Data. *Respir. Res.*, **2024**, *25*, 32.
42. Badaoui, A., De Wergifosse, M., Rondelet, B., Deprez, P.H., Stanciu-Pop, C., Bairy, L., Eucher, P., Delos, M., Ocak, S., Gillain, C., *et al.* Improved Accuracy and Sensitivity in Diagnosis and Staging of Lung Cancer with Systematic and Combined Endobronchial and Endoscopic Ultrasound (EBUS-EUS): Experience from a Tertiary Center. *Cancers*, **2024**, *16*(4), 728.
43. Linga Murthy, M.K., Vinod, S., Koundinya, S., Nagendra Babu, B., Venkataiah, C., Mallikarjuna Rao, Y., Alkhayyat, A., Rana, U. Improving the Accuracy in Lung Cancer Detection Using NN Classifier. *E3S Web Conf.*, **2023**, *391*, 01183.
44. Reddy, N., Khanaa, V. Intelligent Deep Learning Algorithm for Lung Cancer Detection and Classification. *Bull. Electr. Eng. Inform.*, **2023**, *12*, 1747–1754.
45. Shah, A.A., Malik, H.A.M., Muhammad, A., *et al.* Deep Learning Ensemble 2D CNN Approach towards the Detection of Lung Cancer. *Sci. Rep.*, **2023**, *13*, 2987.
46. Srividya, C., Ramasubramanian, K., MadhuBala, M. Federated Learning-Based Diagnostic Approach for Lung Diseases Detection on CT Images. In *Intelligent Computing and*

Communication, Raju, K. S., Senkerik, R., Kumar, T. K., Sellathurai, M., Naresh Kumar, V., Eds., *Lecture Notes in Networks and Systems*, Springer: Singapore, **2025**, pp. 545-554.

47. Nasajpour, M., Pouriyeh, S., Parizi, R.M., Han, M., Mosaiyebzadeh, F., Xie, Y., Liu, L., Batista, D.M. Advances in Application of Federated Machine Learning for Oncology and Cancer Diagnosis. *Information* **2025**, *16*, 487.
48. Cancer Imaging Archive. *Data from The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): A completed reference database of lung nodules on CT scans*. Available online https://www.cancerimagingarchive.net/collection/lidc-idri/ (Access date 11 February 2026).
49. Truhn, D., Arasteh, S.T., Saldanha, O.L., Wetscherek, A., Gessner, F., Meyer, P.T., Lerche, C.W., Hamprecht, F.A. Encrypted Federated Learning for Secure Decentralized Collaboration in Cancer Image Analysis. *Med. Image Anal.*, **2023**, *92*, 103059.
50. Adhikary, S., Dutta, S., Dwivedi, A. D. Secret Learning for Lung Cancer Diagnosis—A Study with Homomorphic Encryption, Texture Analysis and Deep Learning. *Biomed. Phys. Eng. Express*, **2024**, *10*(1), 265102557.
51. Makkar, A., Santosh, K.C. SecureFed: Federated Learning Empowered Medical Imaging Technique to Analyze Lung Abnormalities in Chest X-Rays. *Int. J. Mach. Learn. Cybern.*, **2023**, *14*(8), 2659–2670.
52. Liu, Y., Huang, J., Chen, J.C., Chen, W., Pan, Y., Qiu, J. Predicting Treatment Response in Multicenter Non-Small Cell Lung Cancer Patients Based on Federated Learning. *BMC Cancer*, **2024**, *24*(1), 688.
53. Ali, M., Ahsan, M. M., Tasnim, L., Afrin, S., Biswas, K., Hossain, M., Ahmed, M., Hasan, R., Islam, M., Raman, S. Federated Learning in Healthcare: Model Misconducts, Security, Challenges, Applications, and Future Research Directions—A Systematic Review. *arXiv***2024**, arXiv:2401.00158.
54. Zhang, F., Kreuter, D., Chen, Y., Dittmer, S., Tull, S., Shadbahr, T., BloodCounts! Consortium, Preller, J., Rudd, J. H. F., Aston, J. A. D., Schönlieb, C. B., Gleadall, N., Roberts, M. Recent Methodological Advances in Federated Learning for Healthcare. *Patterns*, **2024**, *5*(6), 101006.
55. Li, M., Xu, P., Hu, J., Tang, Z., Yang, G. From Challenges and Pitfalls to Recommendations and Opportunities: Implementing Federated Learning in Healthcare. *Med. Image Anal.*, **2025**, *101*, 103497.
56. Jin, R., Li, X. Backdoor Attack and Defense in Federated Generative Adversarial Network-Based Medical Image Synthesis. *arXiv***2022**.
57. Karami, M., Ghassemi, F., Kebriaei, H., Azadegan, H. OptiGradTrust: Byzantine-Robust Federated Learning with Multi-Feature Gradient Analysis and Reinforcement Learning-Based Trust Weighting. *arXiv***2025**.
58. Arafat, J., Tasmin, F., Poudel, S., Tareq, A., Haider, I. Beyond Static Knowledge Messengers: Towards Adaptive, Fair, and Scalable Federated Learning for Medical AI. *arXiv***2025**.
59. Alotaibi, K. B., Alam Khan, F., Qawqzeh, Y., Jeon, G., Camacho, D. Performance and Communication Cost of Deep Neural Networks in Federated Learning Environments: An Empirical Study. *Int. J. Interact. Multimed. Artif. Intell.*, **2025**, *9*(4), 6–17.

60. Akshay Kumaar, M., Samiayya, D., Rajinikanth, V., Raj Vincent, P. M. D., Kadry, S. Brain Tumor Classification Using a Pre-Trained Auxiliary Classifying Style-Based Generative Adversarial Network. *Int. J. Interact. Multimed. Artif. Intell.*, **2024**, *8*(6), 101–111.
61. Usharani, C., Selvapandian, A. FedLRes: Enhancing Lung Cancer Detection Using Federated Learning with Convolution Neural Network (ResNet50). *Neural Comput. Appl.* **2025**, *37*, 8273–8284.
62. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N. An Image Is Worth 16 × 16 Words: Transformers for Image Recognition at Scale. *In Proceedings of the 9th International Conference on Learning Representations (ICLR)*, **2021**.
63. Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B. Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows. *In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV*), IEEE: Montreal, QC, Canada, **2021**, pp. 10012–10022.
64. Gopi, S., Mohapatra, P. Learning Cooking Algorithm for Solving Global Optimization Problems. *Sci. Rep.*, **2024**, *14*, 13359.