2025 IEEE International Conference on Signal Processing, Information, Communication and Systems (SPICSCON)
21-22 November 2025, Department of ICE, University of Rajshahi, Rajshahi-6205, Bangladesh

An Explainable Federated EfficientNet–Transformer Model for Lung Cancer Classification from Histopathological CT Images

1st Sinthiya Tabashoum
Faculty of CSE
Patuakhali Sc. and Tech. University
Patuakhali-8602, Bangladesh
sinthiya17@cse.pstu.ac.bd

2nd Aysha Anis Roza
Faculty of CSE
Patuakhali Sc. and Tech. University
Patuakhali-8602, Bangladesh
roza17@cse.pstu.ac.bd

3rd Md. Abdul Masud
Faculty of CSE
Patuakhali Sc. and Tech. University
Patuakhali-8602, Bangladesh
masud@pstu.ac.bd

4th Arjon Golder
Faculty of CSE
Patuakhali Sc. and Tech. University
Patuakhali-8602, Bangladesh
arjonnill07@gmail.com

5th Rakibul Hasan Sezan
Faculty of CSE
Patuakhali Sc. and Tech. University
Patuakhali-8602, Bangladesh
sejan16@cse.pstu.ac.bd

***Abstract*—Lung cancer, a leading cause of cancer-related death, necessitates early and accurate classification. Due to regulatory constraints imposed by HIPAA and GDPR, access to medical data is highly restricted, necessitating the adoption of federated learning approaches to advance sensitive and critical tasks, such as liver cancer classification. However, there remains a significant gap in research concerning the interpretability and explainability of federated medical AI classification models. To address this, we propose an explainable federated deep learning framework that incorporates a hybrid, customized deep neural architecture alongside an advanced aggregation strategy, FedProx. This framework emphasizes model transparency while maintaining data privacy. In This approach the global model achieved almost 98.0% held-out test accuracy over 8 federated rounds, outperforming existing deep learning and federated learning models in both efficiency and accuracy. Furthermore, interpretability analysis using Grad-CAM confirmed that the model reliably attended to neoplastic regions. These results underscore that the combination of federated learning, multi-head attention, and explainable AI provides a strong foundation for advancing digital pathology and establishes a new benchmark by delivering a rare balance of diagnostic accuracy, statistical robustness, and clinical relevance.**

***Index Terms*—Lung cancer classification, Federated learning, EfficientNet-B3, FedProx, Flower framework, Multihead attention mechanism, Non-IID Data, Data Privacy, FedProx Algorithm.**

I. INTRODUCTION

Lung cancer continues to be the foremost cause of cancer-related mortality globally, accounting for approximately 2.2 million newly diagnosed cases and 1.8 million deaths in 2021 alone [1]. Accurate identification and assessment of lung nodules are essential for reliable diagnosis and effective treatment planning. High treatment costs and data privacy issues make diagnosis challenging. Recent advancements in artificial intelligence have significantly transformed the medical field—especially in the detection and diagnosis of lung cancer [1]. As edge devices like medical scanners continue to generate massive amounts of data, conventional centralized deep learning models are increasingly struggling to store the data in a central database and process it effectively [2]. At the same time, concerns about data privacy are rising, and strict regulations like HIPAA and GDPR have made it even more challenging for institutions to exchange personal medical information [3]. This has led to growing interest in Federated Learning (FL), a decentralized approach that enables collaborative model training without exchanging raw data [3]. By keeping data local and sharing only model updates, FL preserves patient confidentiality while still benefiting from large-scale, multi-institutional learning [3].

Although existing federated learning approaches have demonstrated promising accuracy, many are limited by the absence of attention mechanisms, explainable AI frameworks, or support for heterogeneous datasets. Several rely on outdated methodologies or small, low-quality data sources [4]–[6]. In contrast, centralized machine learning introduces additional challenges, including high computational demands, risks to patient privacy, and difficulties in aggregating data across institutions due to strict sharing constraints [7]. These limitations emphasize the necessity for solutions that are not only secure but also interpretable and scalable. To address these limitations, this research introduces a federated deep learning framework for lung cancer classification using chest CT scans. The proposed solution tackles key challenges such as patient data privacy, non-IID data distribution, and

979-8-3315-7316-4/25/$31.00 ©2025 IEEE

limited computational resources by incorporating a Multi-Head Attention mechanism into the learning process. The framework leverages a customized EfficientNet-B3 backbone integrated with Multi-Head Attention for enhanced feature representation. Robustness was further improved through preprocessing steps such as resizing, normalization, and augmentation. Data was distributed across clients using a balanced non-IID partitioning strategy, where local training was conducted independently, and model updates were securely aggregated using FedAvg with FedProx regularization on the Ray platform across multiple rounds. To ensure model interpretability, Grad-CAM was employed, as it provides more precise, fine-grained localization of discriminative regions, thereby improving trust and transparency in clinical decision-making. By combining federated learning with attention mechanisms and explainability, this approach establishes a scalable, privacy-preserving, and clinically relevant solution for real-world healthcare applications.

II. Related Work

Federated Learning (FL) enables multiple devices or institutions to collaboratively train a global model without directly sharing raw data, thereby safeguarding user privacy. Originally introduced through the FedAvg algorithm [8], FL has since evolved with approaches such as FedProx [9], FedNova, SCAFFOLD, MOON, FedOpt, and FedMA, each designed to improve convergence, stability, and performance under heterogeneous (non-IID) data distributions [10]. In medical imaging, and particularly in lung cancer analysis, FL provides a privacy-aware framework by keeping sensitive patient data local. Recent studies demonstrate that FL can achieve classification accuracy comparable to centralized deep learning methods, while maintaining strict privacy constraints. For instance, Mohalder et al [11] reported 95% accuracy on histopathological images using a centralized framework, while other studies achieved 90–93% accuracy for cell-level classification [12]. However, centralized training in medical applications is often limited due to patient confidentiality, data security concerns, and inter-institutional inconsistencies. Given the sensitivity of medical data and the barriers to sharing across institutions, privacy-preserving paradigms such as FL have become indispensable. A growing body of research validates FL's effectiveness in overcoming these challenges, highlighting its potential in lung cancer diagnosis. Nevertheless, the performance of FL models under non-IID conditions remains insufficiently explored, despite real-world datasets being inherently imbalanced and heterogeneous across hospitals and imaging centers. Evaluating FL in such scenarios is therefore essential to ensure robustness and adaptability in clinical practice. Furthermore, this study also considers the computational efficiency and practicality of deploying FL models on mobile and edge devices, enabling distributed training without exposing raw patient data.

III. Methodology

A. Dataset collection and Preparation

This study utilizes a publicly available chest CT scan dataset from Kaggle [13],originally compiled from various online sources for lung cancer research. The dataset consists of 1,000 CT images categorized into four classes: adenocarcinoma (854 images), squamous cell carcinoma (587 images), large cell carcinoma (616 images), and normal (615 images). To standardize the data, the following preprocessing steps were applied:research.

- **Format conversion:** All CT images were converted from DICOM format to standard image formats for ease of use.
- **Resizing:** All images were uniformly resized to **180×180 pixels**.
- **Normalization:** Pixel values were scaled to the range [0,1] by dividing by 255.
- **Data augmentation:** Applied during training to improve generalization, including:
  - Random rotation (±30°)
  - Random shifts and zoom
  - Horizontal and vertical flipping
  - Color jittering (brightness, contrast, saturation)
- **Advanced normalization:** Using standard ImageNet mean and standard deviation values.
- **Tensor conversion:** Ensured compatibility with the PyTorch framework.

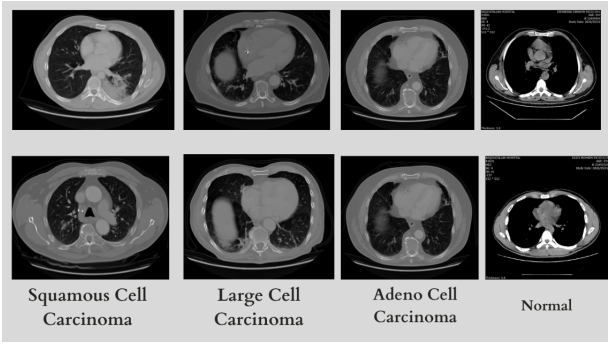

Fig. 1: Sample images of Lung Cancer

B. Proposed Non-IID Partitioning Algorithm

This research simulates site heterogeneity with a *class-shard* (non-Dirichlet) split. After an 80/20 train–test split, we build a $5\times$ offline-augmented training set (original + 4) and partition *that* set. For each class: index $\rightarrow$ shuffle (seed $=$ 2023) $\rightarrow$ shard into $S = 30$. Shards are pooled and assigned round-robin to $N = 5$ clients for multi-class coverage. To induce mild imbalance, we perform $R = 3$ random inter-client transfers (source/destination uniform; transfer size $\sim U[B/2,\, 2B]$, $B = 18$). We then ensure each client has $\geq B$ samples and create a 10% client-local validation.

Algorithm 1 Enhanced Non-IID Partitioning with Class Shards

Input: Dataset D with C classes, number of clients N, batch size B, shard size S

Output: Client indices $client\_indices$ with realistic non-IID partitions

Initialize a dictionary $class\_indices$ with C empty lists **for** *each* $(index, label)$ *in* D **do**
  Append $index$ to $class\_indices[label]$

for *each list in* $class\_indices$ **do**
  Shuffle the list and split into shards of size S

Initialize $client\_indices$ with N empty lists **for** *each class* c *in* C **do**
  Distribute shards of class c across clients using round-robin Randomly assign extra shards to simulate imbalance

$min\_size \leftarrow \min(\{|list| \mid list \in client\_indices\})$ **if** $min\_size < B$ **then**
  Rebalance or adjust shard size S to satisfy batch requirements

return $client\_indices$

This ensured each client received diverse yet slightly imbalanced data, reflecting practical healthcare scenarios while preserving robustness and generalization in the federated learning setup.

C. Proposed Custom Model Architecture

In this study, a custom deep learning model was developed depicted in Figure 3, using EfficientNetB3 as the base, known for its strong feature extraction and efficiency, which is important for federated learning. The original EfficientNetB3 was modified by replacing its final layer with a fully connected layer that outputs 1024 features. Features are then processed by a Multi-Head Self-Attention (MHA) block (Fig 2) with embed_dim = 1024 and 8 heads ($d_k = 128$), implemented as scaled dot-product attention [14]. The 1024-D backbone feature is treated as a single

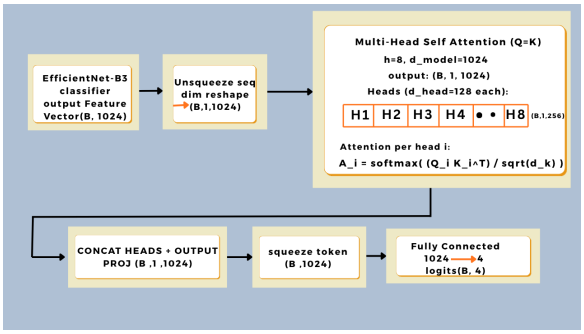

Fig. 2: Custom Multihead Attention Mechanism

token ($B \times 1 \times 1024$; $Q = K = V$), with a dropout ($p = 0.5$) applied beforehand. The attention output feeds a fully connected layer ($1024 \rightarrow 4$) to predict adenocarcinoma, squamous cell carcinoma, large cell carcinoma, or normal tissue. This architecture combines EfficientNetB3's powerful feature extraction capability with attention mechanisms ability to capture complex image details and underlying patterns while managing challenges like limited computational resources. All the above capabilities making federated learning, well-suited for medical imaging tasks distributed across multiple locations.

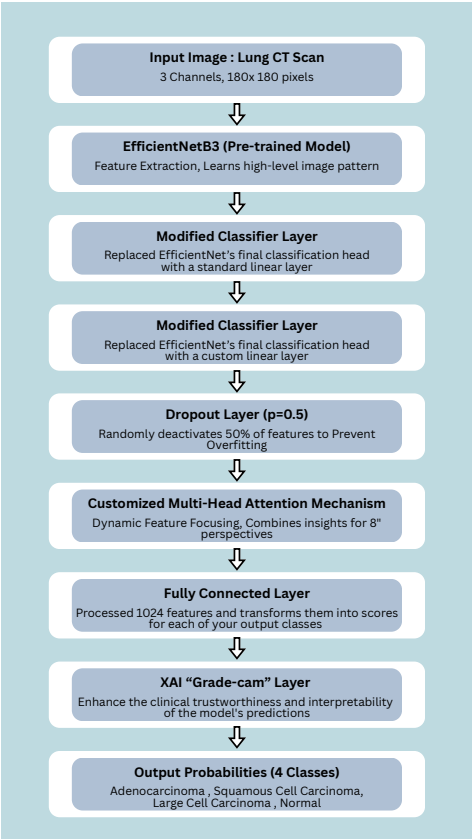

Fig. 3: Proposed model architecture with Custom EfficientNetB3

D. Explainability: Grad-CAM

To ensure decisions are clinically grounded—not just accurate—we perform post-hoc explanations with Grad-CAM on the final global model (after the last federated round). Because Grad-CAM requires spatial feature maps, we target the last convolutional layer of EfficientNet-B3. For class c with activation maps $A_k \in \mathbb{R}^{H \times W}$ and logit y^c, we compute channel weights and the heatmap . The map is normalized to $[0, 1]$, bilinearly upsampled to 224×224, and alpha-blended ($\alpha = 0.35$) onto the de-normalized CT image. Implementation uses forward/backward hooks and single-image inference to control memory. Figure 4 shows four correctly classified cases where highlighted regions align with suspected neoplastic areas, complementing our

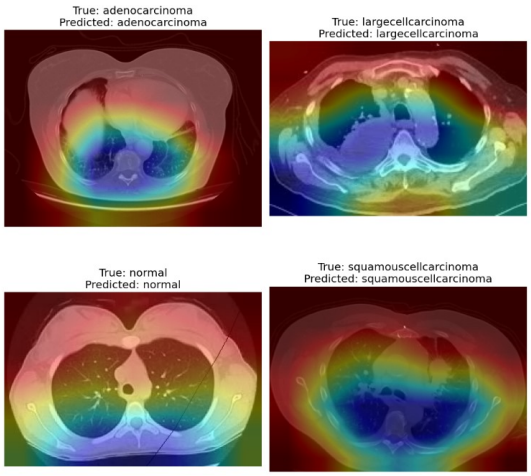

Fig. 4: Grad-CAMvisualization of correctly classifiedCT imagesfor thefour lungcancercategories

quantitative gains and supporting interpretability under a privacy-preserving, federated setup.

E. Training Process

The training process and federated learning implementation were carefully designed with tuned hyperparameters to ensure performance and stability. The model is optimized for deployment in resource-constrained environments, such as healthcare edge devices. Experiments on Google Colab's T4 GPU required only 8GB RAM and 10GB GPU memory, demonstrating robust performance without reliance on costly high-end infrastructure.

TABLE I: Evaluation metrics over training rounds

Round	Precision	Recall	F1-Score	Accuracy	Centralized Loss	Distributed Loss	AUC	Sensitivity
0	0.182	0.293	0.171	0.3205	4.129	N/A	0.587	0.310
1	0.893	0.879	0.878	0.7413	3.259	2.068	0.980	0.872
2	0.975	0.975	0.975	0.8958	1.650	1.148	0.998	0.966
3	0.980	0.980	0.980	0.9459	1.705	1.081	0.998	0.979
4	0.987	0.987	0.987	0.8996	0.664	0.820	0.996	0.975
5	0.987	0.987	0.987	0.9575	0.563	0.436	0.999	0.974
6	0.993	0.993	0.993	0.9421	0.751	0.603	1.000	0.985
7	0.993	0.993	0.993	0.9537	0.360	0.530	1.000	0.974
8	0.993	0.993	0.993	0.9791	0.219	0.498	1.000	0.982

F. Proposed Federated Deep Learning Approach

The proposed approach adopts a horizontal FL framework ((Fig 5)) with a central server and five clients; raw data remain on-device to preserve privacy. Each client performs class-balanced non-IID partitioning and local augmentation, then trains a customized EfficientNet-B3 with Multi-Head Attention [15], [16]. Clients send model updates—not data—via Ray [17]; the server aggregates with FedAvg [8], while FedProx (=0.01) regularizes local objectives to address heterogeneity [9]. Training proceeds for 8 rounds with 60% client participation per round, and the global model achieves 98.0% held-out test accuracy, yielding a scalable, privacy-preserving, and computationally efficient solution for lung-cancer CT classification.

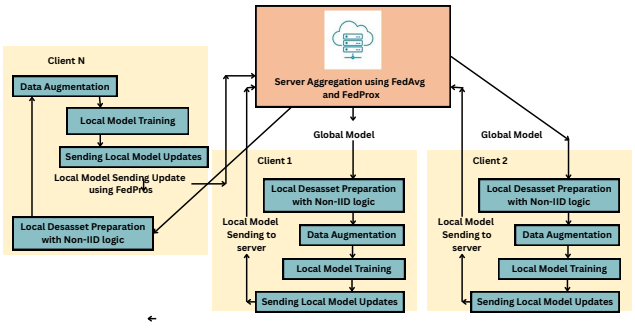

Fig. 5: Proposed federated deep learning approach aggregate FedAvg with Fedprox strategy

IV. RESULTS AND DISCUSSION

A. The Performance of the Model

The proposed federated deep learning model for lung cancer classification was rigorously evaluated over eight training rounds.As illustrated in Figure 6.

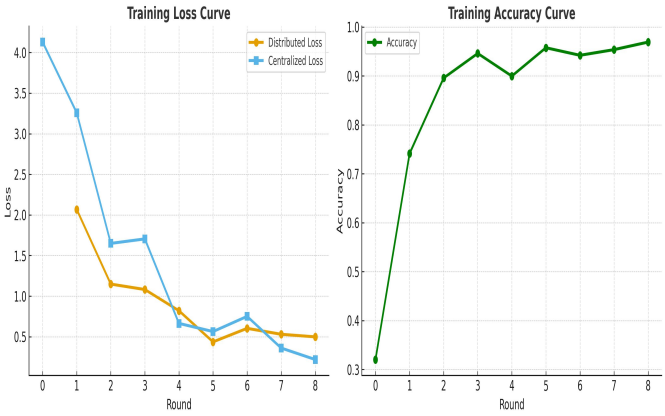

Fig. 6: Model accuracy and loss over rounds

The model's centralized accuracy improved significantly from 32.05% in the initial round to 97.91% by Round 8. This indicates rapid early learning and strong convergence, particularly under non-IID conditions.Correspondingly, centralized loss dropped sharply from 4.129 to 0.219, while distributed loss decreased from 2.068 to 0.498 Figure 5, confirming effective optimization using FedAvg with FedProx regularization.Confusion matrices Figure 7 further support this progress. Initial rounds showed high misclassification—especially into a dominant class—yet accuracy across all four lung cancer classes (adenocarcinoma, large cell

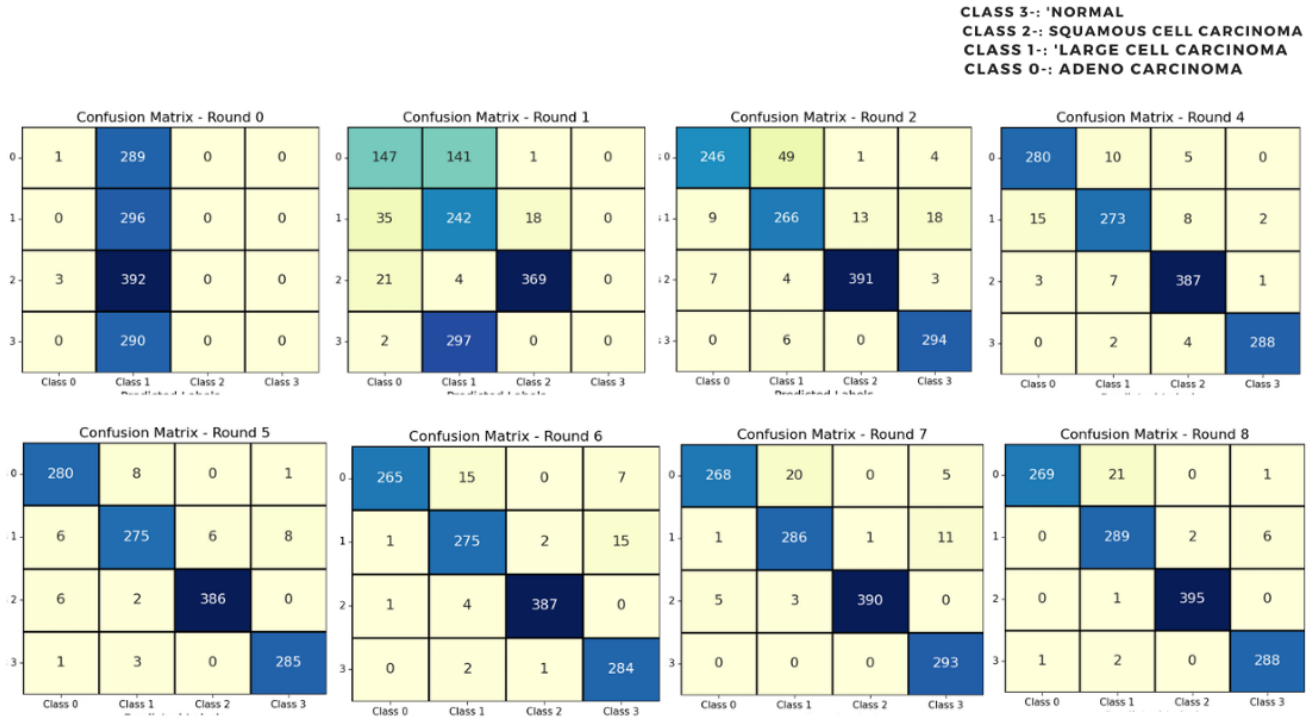

Fig. 7: Confusion matrices over multiple rounds

carcinoma, squamous cell carcinoma, and normal) steadily improved. By the final round, the model showed minimal misclassification and high class-specific accuracy.

TABLE II: Evaluation metrics over training rounds

Round	Precision	Recall	F1-Score	Accuracy	Centralized Loss	Distributed Loss	AUC	Sensitivity
0	0.182	0.293	0.171	0.3205	4.129	N/A	0.587	0.310
1	0.893	0.879	0.878	0.7413	3.259	2.068	0.980	0.872
2	0.975	0.975	0.975	0.8958	1.650	1.148	0.998	0.966
3	0.980	0.980	0.980	0.9459	1.705	1.081	0.998	0.979
4	0.987	0.987	0.987	0.8996	0.664	0.820	0.996	0.975
5	0.987	0.987	0.987	0.9575	0.563	0.436	0.999	0.974
6	0.993	0.993	0.993	0.9421	0.751	0.603	1.000	0.985
7	0.993	0.993	0.993	0.9537	0.360	0.530	1.000	0.974
8	0.993	0.993	0.993	0.9791	0.219	0.498	1.000	0.982

The model demonstrated increasing performance in precision, recall, and F1-score—starting at 0.182, 0.293, and 0.171 in Round 0, and reaching 0.993 across all three metrics by Round 8 according to Table II.The F1-score, a metric that balances both precision and recall, increased significantly from 0.171 in the initial round to 0.993 by the final round, highlighting a substantial improvement in the model's overall classification performance as detailed in Table III.This performance gain is attributed to the integration of a Multi-Head Attention mechanism on top of EfficientNetB3, which enables the model to focus on important spatial and contextual features.

B. The Performance Comparison and Discussion

Our federated model—EfficientNet-B3 with Multi-Head Attention—matches or exceeds conventional ML and contemporary FL baselines (Table IV). In an ablation isolating FedProx under identical settings (5 clients, 8 rounds, 60% participation, non-IID class shards, AdamW), FedAvg without FedProx ($\mu = 0$) reached **95.0%** test accuracy, while FedAvg + FedProx ($\mu = 0.01$) achieved **98.0%** (+3.0 pp), indicating greater robustness to client drift and data heterogeneity.

TABLE III: **Comparison of Different Deep Learning Models and Federated Aggregation Strategies for Lung Cancer Classification**

Deep Learning Model	Aggregation Strategy	Accuracy
EfficientNetB0	FeddropoutAvg	91.0%
EfficientNetB0	FedProx	89.0%
EfficientNetB0	FedDyne	82.0%
EfficientNetB0	FedOpt	95.0%
MobileNetV2	FedOpt	83.0%
MobileNetV3	FedProx	89.62%
MobileNetV3	FedOpt	95.43%
Proposed Model	**FedAvg**	**95.0%**
Proposed Model	**FedAvg with FedProx regularization**	**98.0%**

TABLE IV: **Comparison of different models for Lung Cancer Classification**

Study	Accuracy	Technique
Muntasir Mamun [6]	92.0%	Deep learning based approach with CNN (ResNet-50, Inception V3)
CHAMAK SAHA [5]	91.5%	Lightweight attention-based CNN (LGAM) in Federated learning setup (FedAvg)
Umamaheswa-ran Sub-ashchan-drabose [18]	89.63%	Ensemble Federated Learning with Data calibration and feature-set map- ping
Smitirekha Behuria [19]	89.0%	Federated learning including fine-tuning and transfer learning
Proposed Model	**98.0%**	**Customized EfficientNetB3 with Multi-Head Attention mecha-nism, FedAvg with FedProx reg-ularization**

Using the Flower framework [20] and FedProx algorithm, our proposed federated deep learning model ensures data privacy by decentralizing data management. The proposed model achieved a centralized accuracy almost 98.0%, as shown in Table IV, demonstrsg superior performance when compared to previously published methods. For example, MMuntasir Mamun's ResNet-50/Inception V3 (92.0%) [9], CHAMAK SAHA's LGAM (91.5%) [8], and other federated techniques (89–92%) [22], [23], achieved lower accuracy.

V. Conclusion

This study introduces a federated deep learning framework for lung cancer classification from CT scan images, addressing data privacy challenges inherent in centralized processing. The proposed model integrates EfficientNetB3 with a multi-head attention mechanism to enhance feature extraction and classification across four lung cancer types. Implemented with the Flower framework using the FedAvg algorithm, it achieved 98% centralized accuracy within eight training rounds while operating efficiently on modest hardware (8GB RAM, 10GB GPU). To ensure transparency and clinical trust, explainable AI methods such as Grad-CAM was employed to highlight decision-relevant regions, validating the model's focus on pathological areas. This framework demonstrates that federated learning, combined with attention mechanisms and explainability, offers a scalable, privacy-preserving, and effective solution for AI-driven medical diagnostics, with future work extending toward histopathological data and broader client environments.

References

[1] D. D. Miller and E. W. Brown, "Artificial intelligence in medical practice: the question to the answer?," *The American journal of medicine*, vol. 131, no. 2, pp. 129–133, 2018.

[2] S. K. Zhou, H. Greenspan, C. Davatzikos, J. S. Duncan, B. Van Ginneken, A. Madabhushi, J. L. Prince, D. Rueckert, and R. M. Summers, "A review of deep learning in medical imaging: Imaging traits, technology trends, case studies with progress highlights, and future promises," *Proceedings of the IEEE*, vol. 109, no. 5, pp. 820–838, 2021.

[3] R. Sharma, J. Miller, P. Iyer, and C. James, "Federated learning in healthcare: Privacy-preserving ai for secure medical data analysis," *ResearchGate. Viitattu*, vol. 25, p. 2025, 2024.

[4] A. A. Nafea, M. S. Ibrahim, M. M. Shwaysh, K. Abdul-Kadhim, H. R. Almamoori, and M. M. AL-Ani, "A deep learning algorithm for lung cancer detection using efficientnet-b3," *Wasit Journal of Computer and Mathematics Science*, vol. 2, no. 4, pp. 68–76, 2023.

[5] C. Saha, S. Saha, M. A. Rahman, M. H. Milu, H. Higa, M. A. Rashid, and N. Ahmed, "Lung-attnet: An attention mechanism based cnn architecture for lung cancer detection with federated learning," *IEEE Access*, 2025.

[6] M. Mamun, M. I. Mahmud, M. Meherin, and A. Abdelgawad, "Lcd-ctcnn: Lung cancer diagnosis of ct scan images using cnn based model," in *2023 10th International Conference on Signal Processing and Integrated Networks (SPIN)*, pp. 205–212, IEEE, 2023.

[7] Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated machine learning: Concept and applications," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 10, no. 2, pp. 1–19, 2019.

[8] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Artificial intelligence and statistics*, pp. 1273–1282, PMLR, 2017.

[9] A. Das and D. Saha, "Fedprox-based federated transfer learning for efficient model personalization in healthcare," pp. 1–6, 01 2025.

[10] M. Ye, X. Fang, B. Du, P. C. Yuen, and D. Tao, "Heterogeneous federated learning: State-of-the-art and research challenges," *ACM Computing Surveys*, vol. 56, no. 3, pp. 1–44, 2023.

[11] R. Mohalder, K. Hossain, J. Sarkar, L. Paul, M. Raihan, and K. Talukder, *Lung Cancer Detection from Histopathological Images Using Deep Learning*, pp. 201–212. 06 2023.

[12] S. Wang, T. Wang, L. Yang, D. Yang, J. Fujimoto, F. Yi, X. Luo, Y. Yang, B. Yao, S. Lin, C. Moran, N. Kalhor, A. Weissferdt, J. Minna, Y. Xie, I. Wistuba, Y. Mao, and G. Xiao, "Convpath: A software tool for lung adenocarcinoma digital pathological image analysis aided by a convolutional neural network," *EBioMedicine*, vol. 50, 11 2019.

[13] M. Hany, "Chest ct-scan images dataset." https://www.kaggle.com/datasets/mohamedhanyyy/chest-ctscan-images, 2020. Accessed: July 15, 2025.

[14] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proceedings of the 31st International Conference on Neural Information Processing Systems (NeurIPS)*, pp. 5998–6008, 2017.

[15] F. Adnan, M. Awan, A. Mahmoud, H. Nobanee, A. Yasin, and A. Zain, "Efficientnetb3-adaptive augmented deep learning (aadl) for multi-class plant disease classification," *IEEE Access*, vol. 11, pp. 85426 – 85440, 08 2023.

[16] A. Wu and Y.-W. Kwon, "Enhancing recommendation capabilities using multi-head attention-based federated knowledge distillation," *IEEE Access*, vol. PP, pp. 1–1, 01 2023.

[17] P. Moritz, R. Nishihara, S. Wang, A. Tumanov, R. Liaw, E. Liang, W. Paul, M. I. Jordan, and I. Stoica, "Ray: A distributed framework for emerging ai applications," *ArXiv*, vol. abs/1712.05889, 2017.

[18] U. Subashchandrabose, R. John, U. Anbazhagu, V. V. Kumar, and M. T R, "Ensemble federated learning approach for diagnostics of multi-order lung cancer," *Diagnostics*, vol. 13, p. 3053, 09 2023.

[19] S. Behuria, S. Swain, A. Bandyopadhyay, S. M. Turjya, and M. Gourisaria, "Federated learning approach in healthcare ecosystems for efficient lung cancer classification: Insights from model generic training to fine-tuning and transfer learning," *Procedia Computer Science*, vol. 259, pp. 279–290, 01 2025.

[20] Y. Liu, J.-J. Peng, J. Kang, A. M. Iliyasu, D. Niyato, and A. El-latif, "A secure federated learning framework for 5g networks," *IEEE Wireless Communications*, vol. 27, pp. 24–31, 2020.