XPA-FedCBR: A Federated Case-Based Reasoning System for Explainable and Privacy-Aware Lung Cancer Diagnosis

Devyani Rawat[1], Shuchi Bhadula[2], Sachin Sharma[3]
Department of Computer Science and Engineering, Graphic Era Deemed to be University, Dehradun, India[1, 2]
Amity School of Engineering and Technology, Amity University Punjab, Mohali, India[3]

***Abstract*—Lung cancer is among the deadliest cancers worldwide, largely because it is usually caught too late. Building AI tools for earlier, stage-aware diagnosis is hard in practice: patient scans are scattered across hospitals that cannot share them freely, and clinicians are reluctant to trust models that cannot explain their reasoning. This study presents XPA-FedCBR, a framework that addresses both concerns. It combines Federated Learning, which trains a shared model across institutions without moving raw data, with Case-Based Reasoning, which supports each prediction by retrieving clinically similar past cases and, for uncertain predictions, using them to refine the decision. We evaluate it on the LIDC-IDRI dataset (roughly 1,018 CT scans from 1,010 patients), split across five simulated institutions under non-IID conditions, with annotations mapped to stages I–IV. Over four independent runs, the proposed model reaches a macro-F1 of 0.915, significantly ahead of FedAvg without CBR (0.875), a centralized CNN (0.895), and a standalone CNN (0.844), with its advantage largest under domain shift. Its case-based explanations, evaluated quantitatively and against gradient-based saliency, give clinicians evidence they can directly inspect.**

***Keywords*—Case-based reasoning; federated learning; explainable AI; lung cancer; clinical decision support; medical imaging**

I. Introduction

A. Clinical Background

One of the most common and deadly types of cancer in the world, lung cancer accounts for a large percentage of cancer-related deaths. The high fatality rate is primarily attributed to the late-stage detection of the disease, as early-stage lung cancer often presents minimal or no noticeable symptoms[1]. As a result, a large number of patients are diagnosed only after the cancer has progressed to an advanced stage, where treatment options become limited, and survival rates decline substantially.

In order to improve patient survival and treatment outcomes, early and precise identification of lung cancer is essential. When detected at an early stage, lung cancer can be managed more effectively through surgical intervention, targeted therapy, or radiotherapy, leading to significantly higher survival rates[2]. Therefore, timely screening and reliable diagnostic tools are essential for reducing mortality and improving the overall quality of patient care.

Non-Small Cell Lung Cancer (NSCLC), which accounts for most lung cancer cases, is staged from I to IV according to tumor size, lymph node involvement, and metastasis. The stage matters enormously for outcome: patients caught at Stage I or II have far better survival odds, while Stage III and IV carry a poor prognosis [3]. Because the stage drives the choice of treatment, a diagnosis that simply flags cancer as present or absent is not enough; what clinicians need is a system that can tell the stages apart.

Despite advances in medical imaging such as computed tomography (CT), lung cancer screening still has real limitations. Much of it depends on manual interpretation by radiologists, which is time-consuming and varies from one reader to the next [4]. Subtle nodules are easy to miss, especially early on, and both false positives and false negatives take their toll, the former sending patients into unnecessary follow-ups and anxiety, the latter delaying diagnoses that matter. These gaps are why the field needs diagnostic tools that are not just accurate but consistent and clinically trustworthy, and why computational methods have drawn so much attention as a way to close them.

B. Challenges in AI-Based Lung Cancer Detection

AI has shown real promise in lung cancer analysis, but several obstacles keep it from being adopted in actual clinical practice. One is data scarcity and institutional silos. Well-annotated imaging data is hard to come by and tends to sit locked inside individual hospitals, which, for reasons of patient confidentiality and limited infrastructure, cannot share it freely. A model trained in one institution therefore rarely sees the range of patients it would need to generalize well.

A second obstacle is privacy and regulation. CT scans and clinical records are among the most sensitive data a hospital holds, yet conventional centralized learning depends on pooling everything in one place, an approach that can run afoul of rules like HIPAA and GDPR and leaves institutions understandably wary of any system built around direct data sharing [7].

The third is the opacity of deep learning itself. These models can be highly accurate, but they say little about *how* they reach a decision, and a clinician who cannot follow that reasoning has no way to check an AI-assisted diagnosis in something as consequential as cancer staging. Taken together, these problems point to a clear need: systems that can learn from data spread across institutions, respect privacy law, and still explain themselves in terms a clinician can act on.

C. Motivation

The challenges outlined above- fragmented data, privacy constraints, and the opacity of deep models- are rarely addressed together, yet in clinical practice they arise together. A system that preserves privacy but cannot explain its decisions will not earn clinician trust; one that explains its reasoning but requires centralized data cannot be deployed across institutions. This motivates a combined approach. Federated learning offers a way to train across distributed hospitals without moving sensitive data, directly addressing the privacy and data-silo problem. Case-based reasoning offers a complementary strength: by grounding each prediction in similar, previously confirmed cases, it provides the transparent, evidence-based justification clinicians rely on when making decisions. Pairing the two is therefore natural; federated learning supplies privacy-preserving predictive capability, while case-based reasoning supplies the interpretability that turns a prediction into a clinically actionable, trustworthy recommendation. Combining the two, rather than using either one alone, is the main idea behind this work.

Despite these advances, existing approaches address the key challenges only in isolation. Federated learning methods for lung cancer have focused primarily on improving accuracy and preserving privacy, but they largely operate as black boxes and offer little in the way of clinician-facing explanation. Case-based reasoning, by contrast, provides transparent, evidence-based justification that aligns well with clinical decision-making, yet it has rarely been integrated with modern deep or distributed learning and does not, on its own, address privacy or scale to high-dimensional imaging data. Furthermore, most existing systems are limited to binary malignant-versus-benign detection rather than clinically meaningful stage-aware classification. As a result, no unified framework currently delivers privacy-preserving collaborative learning, stage-level diagnosis, and clinician-understandable explanations together. This gap motivates the hybrid framework proposed in this work, which couples federated learning with case-based reasoning to achieve all three within a single architecture.

D. Contributions

The main contributions of this work are summarized as follows:

- A hybrid Case-Based Reasoning and Federated Learning framework is proposed for lung cancer detection, enabling collaborative model training while preserving patient data privacy.
- Introduces stage-aware lung cancer classification for clinically meaningful diagnosis.
- An explainable clinical decision-support layer is introduced, which aligns AI predictions with clinician reasoning by retrieving and presenting similar historical patient cases.
- Robustness across decentralized clinical data sources is demonstrated by evaluating the proposed framework under a simulated multi-institution federated learning environment.
- A comprehensive comparative evaluation is performed against centralized learning and non-explainable baseline models to confirm the proposed framework's efficacy.

II. Related Work

A. Lung Cancer Detection Using AI

Early efforts in lung cancer diagnosis using artificial intelligence were built around traditional computer-aided diagnosis (CAD) systems. In these systems, handcrafted features such as shape, texture, and intensity of lung nodules were manually extracted from CT images, followed by conventional machine learning classifiers like support vector machines and decision trees for detection and classification. Such CAD systems were designed to assist radiologists and reduce workload, but they typically depended heavily on engineered features and showed limited adaptability across diverse clinical data sources[5]. A hybrid transfer learning framework, VER-Net (2024), combined VGG19, EfficientNetB0, and ResNet101 to classify lung cancer from CT images. Using the Chest CT-Scan Images Dataset (Kaggle), the proposed model achieved 91.0% accuracy, 92.0% precision, 91.0% recall, and an F1-score of 91.3%, highlighting the effectiveness of hybrid deep learning for automated lung cancer diagnosis[6].

B. Federated Learning in Healthcare

A proximal-guided hybrid federated learning framework for pneumonia diagnosis that combines FedProx with parameter-efficient adaptive intelligence to address data heterogeneity across hospitals. The model was evaluated on Chest X-ray datasets collected from multiple healthcare institutions and achieved 98.7% accuracy, demonstrating improved robustness and communication efficiency compared with conventional federated learning methods[7].

One of the central motivations for applying FL in healthcare is its ability to address privacy and regulatory constraints such as HIPAA and GDPR by keeping medical data within the originating institution while still enabling collaborative model development. This decentralization reduces the risk of sensitive data exposure and aligns with legal and ethical requirements for handling patient information[8].

C. Case-Based Reasoning in Clinical Systems

CBR is a problem-solving approach that draws on solutions to previously encountered cases to inform decisions about new, similar situations. In clinical settings, CBR has been applied to diagnostic support by retrieving past patient cases with similar clinical features and outcomes to guide current decision-making. This reasoning paradigm aligns closely with the way clinicians think, often referring to past patient histories, symptoms, and outcomes when diagnosing new patients [9]. A hybrid knowledge modeling framework incorporating Case-Based Reasoning (CBR) and rule-based reasoning was proposed to recommend treatment strategies for patients with chronic kidney disease–mineral and bone disorder (CKD-MBD). The system was validated using electronic health records and clinical guidelines, providing treatment recommendations that closely matched expert clinical decisions while improving consistency in decision-making [10].

An enhanced Case-Based Reasoning (CBR) framework integrated with a Rule-Based System (RBS) was developed to support medical diagnosis by retrieving similar historical patient cases and refining diagnostic decisions using expert rules. The system was evaluated using a medical diagnosis dataset and achieved an overall diagnostic accuracy of 96.8%, demonstrating the effectiveness of combining CBR with rule-based reasoning for clinical decision support[11]. An explainable Case-Based Reasoning (XCBR) framework was developed for lung cancer diagnosis by integrating Naïve Bayes, Multilayer Perceptron (MLP), SHAP, and Harris Hawks Optimization for feature selection and explanation. The model was evaluated on two public lung cancer datasets and achieved 94.47% and 100% classification accuracy, demonstrating the effectiveness of combining CBR with explainable AI for clinical decision support[12].

D. Research Gap

Prior work has advanced federated and case-based approaches for lung cancer diagnosis, but along separate tracks. In the federated setting, Liu et al. [14] applied federated learning across multiple centres to predict treatment response in non-small cell lung cancer, and Heidari et al. [15] combined federated learning with blockchain for CT-based lung cancer detection; both demonstrate that collaborative training without data sharing is feasible, yet neither incorporates a mechanism to explain individual predictions to clinicians. Case-based reasoning offers precisely this missing interpretability; Pradeep et al. [13] show that retrieving and presenting similar past cases yields transparent, clinically aligned justification, but such systems are typically built on manually defined features and have rarely been coupled with deep or distributed learning, leaving them unable to scale to high-dimensional imaging or to operate under privacy constraints. A further limitation is that most of these systems target binary malignant-versus-benign detection rather than stage-level diagnosis, which is what ultimately guides treatment decisions. Consequently, no existing framework simultaneously provides privacy-preserving collaborative learning, stage-aware classification, and case-based explanations suited to clinical use. The proposed work addresses this gap by unifying all three within a single architecture.

III. Proposed Methodology

A. System Overview

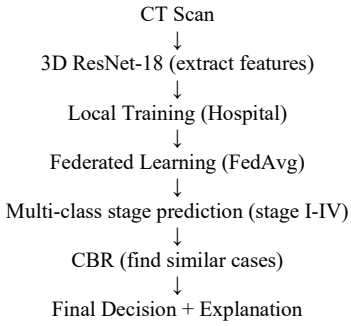

Fig. 1. Proposed hybrid CBR-FL framework.

The proposed intelligent digital clinical diagnostic system is designed to support lung cancer diagnosis while preserving patient privacy and providing explainable clinical decisions. Fig. 1 shows that the proposed framework processes CT scan images using a CNN-based feature extraction model followed by federated learning for privacy-preserving multi-class lung cancer stage prediction. The predicted stage is further supported through a Case-Based Reasoning module that retrieves clinically similar historical cases to provide interpretable diagnostic explanations.

The overall architecture follows a distributed learning framework in which multiple healthcare institutions act as local clients, and a central server coordinates the global learning process. Each local client trains a diagnostic model using its own patient data, ensuring that sensitive clinical information does not leave the institution. The CNN-based model is extended to perform multi-class classification, where each class corresponds to a specific stage of lung cancer. Table I shows the detailed stages of Non-Small Cell Lung Cancer.

TABLE I. Detailed Stages of Non-Small Cell Lung Cancer

Stage 1	Cancer is only located in the lungs.
Stage 2	Cancer is located in the lungs and may be in the lymph nodes near the lungs.
Stage 3	Cancer is located in the lungs and lymph nodes in the middle of the chest.
Stage 3A	Cancer is extensive but localized to one side of the lung.
Stage 3B	Cancer has spread to lymph nodes on both sides of the lungs.
Stage 4	Cancer has spread to both lungs or to another part of the body. This stage is considered the most advanced.

During training, only the locally learned model parameters are shared with the global server, which aggregates them to update a global model using a federated learning strategy. This global model is then redistributed to the local clients for further refinement. In parallel, a case-based reasoning module maintains a repository of historical patient cases and retrieves clinically similar cases to support transparent and interpretable decision-making. By combining federated learning with case-based reasoning, the system enables accurate lung cancer diagnosis while ensuring data privacy and clinician-understandable explanations. Fig. 2 shows the proposed hybrid case-based learning and federated learning framework for privacy-preserving lung cancer stage prediction using CT scan images.

B. Local Client-Side Processing and Feature Extraction

In the proposed system, lung CT scan images are collected at individual hospitals and processed locally to ensure patient data privacy. Each CT scan is first passed through a 3D ResNet-18 convolutional neural network (CNN), which automatically extracts relevant features such as lung nodules, texture patterns, and structural abnormalities. These features represent important visual characteristics required for lung cancer detection. The extracted features are then used to train a local diagnostic model within the hospital environment. At no point is raw patient data shared outside the institution.

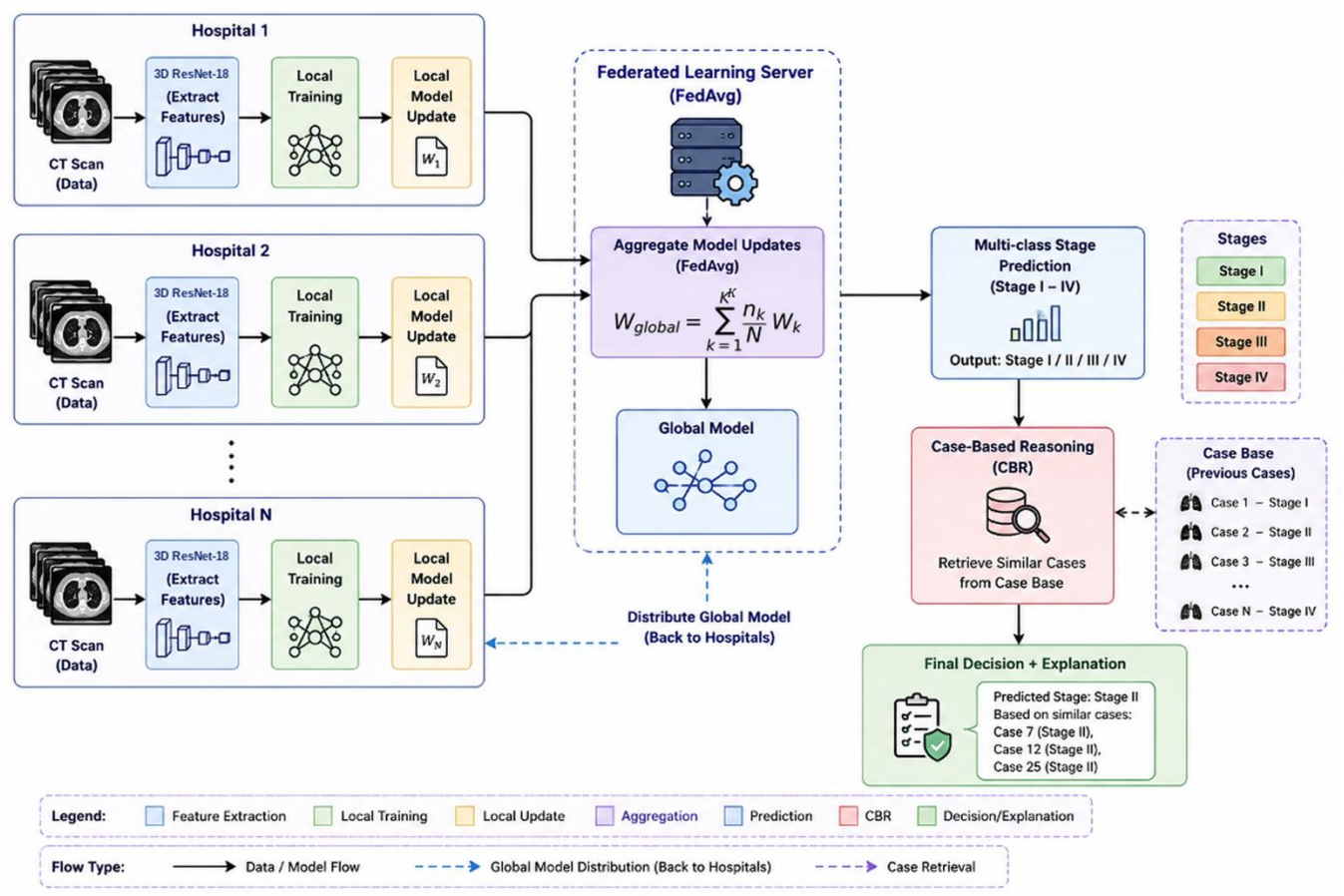

Fig. 2. Proposed hybrid case-based reasoning and federated learning framework for privacy-preserving lung cancer stage prediction using CT scan images.

C. Federated Learning for Collaborative Model Training

To enable collaborative learning across multiple clinical centers while preserving patient privacy, a federated learning framework is adopted in the proposed system. Instead of sharing raw CT scan data, each participating hospital trains the diagnostic model locally using its own dataset. This ensures that sensitive patient information remains within the institution and complies with privacy regulations.

After local training, only the updated model parameters are transmitted to a central server. The server performs secure aggregation of these updates using the Federated Averaging (FedAvg) algorithm to generate a global model that captures collective knowledge from all participating centers. The updated global model is then redistributed to the local clients for further training in subsequent rounds.

Since clinical data from different hospitals may vary due to differences in imaging equipment, patient demographics, and acquisition protocols, the federated framework is designed to handle data heterogeneity. Through iterative aggregation across multiple training rounds, the global model becomes more robust and generalized, learning patterns that are consistent across diverse clinical environments. Furthermore, the predictions generated by the global model are supported by a case-based reasoning mechanism, which retrieves clinically similar cases and uses their confirmed stage labels both to explain and, for low-confidence predictions, to refine the final decision.

In federated learning, suppose there are K participating clinical centers. Each hospital k trains a local model on its dataset containing n_k samples. After local training, the model parameters w_k are transmitted to the central server.

The global model parameters at communication round t+1 are computed using weighted averaging as follows:

$$w_{t+1} = \Sigma_{k=1}^{K} (n_k / N) \cdot w_k^t \quad (1)$$

where,

w_k^t = local model parameters from hospital k at round t

n_k = number of samples at hospital k

$N = \Sigma_{k=1}^{K} n_k$ = total number of samples across all hospitals

K = total number of participating hospitals

This weighted aggregation ensures that hospitals with larger datasets contribute proportionally more to the global model update.

D. Case-Based Reasoning Module

The Case-Based Reasoning (CBR) module is incorporated to provide interpretability and support clinical decision-making by leveraging previously diagnosed cases. In this approach, each patient case is represented using a combination of imaging features extracted from CT scans and relevant clinical

information, if available. This structured representation allows the system to compare new cases with historical records effectively.

To identify similar cases, a similarity measurement technique is applied. In this work, cosine similarity is used to compute the closeness between the feature vector of the new patient and those stored in the case database. Based on this similarity score, the top-k most relevant cases are retrieved.

Once similar cases are identified, the system analyses their confirmed diagnoses. These retrieved cases serve two complementary purposes: they are presented alongside the model's prediction to provide context and justification, and, for low-confidence predictions, their stage labels contribute to the final decision through the fusion mechanism described in Section III-E. This process enables explanations that are intuitive and aligned with clinical reasoning, while allowing similar historical cases to reinforce or correct uncertain predictions, thereby improving the transparency, trustworthiness, and reliability of the diagnostic system.

E. Hybrid Integration Strategy

The proposed framework integrates federated learning (FL) and case-based reasoning (CBR) to combine accurate prediction with interpretable, evidence-based decision support. Rather than treating the two components independently, the system couples them so that CBR both explains and, where appropriate, refines the federated model's output.

For a query case with deep feature vector q (extracted by the shared 3D ResNet-18 backbone), the global federated model produces a probability distribution over the four lung cancer stages, $p = (p_1, p_2, p_3, p_4)$, and a corresponding confidence $\hat{c} = \max_j p_j$. This distribution serves as the initial diagnostic indicator, which is then passed to the CBR module.

1) Case retrieval: To identify clinically similar precedents, the CBR module compares **q** against every stored case *i* (with feature vector $\mathbf{c}_i$) using cosine similarity:

$$\mathrm{sim}(\mathbf{q}, \mathbf{c}_i) = (\mathbf{q} \cdot \mathbf{c}_i) / (\|\mathbf{q}\| \, \|\mathbf{c}_i\|). \quad (2)$$

The top-*k* cases with the highest similarity are retrieved and denoted $\mathcal{N}$. In this work k = 5.

2) Neighbour label evidence: The retrieved cases contribute a similarity-weighted label distribution, in which each neighbour supports its own confirmed stage in proportion to how similar it is to the query:

$$v_j = \left(\Sigma_{i \in \mathcal{N} : y_i = j} \, \mathrm{sim}(\mathbf{q}, \mathbf{c}_i) \right) / \left(\Sigma_{i \in \mathcal{N}} \, \mathrm{sim}(\mathbf{q}, \mathbf{c}_i) \right), \quad \text{for each stage } j \quad (3)$$

where, y_i is the confirmed stage label of case *i* and $\mathbf{v} = (v_1, v_2, v_3, v_4)$.

3) Confidence-gated decision fusion: In the decision-fusion stage, the quantitative prediction from the FL model is combined with the qualitative evidence from retrieved cases through a confidence-gated rule:

$$\mathbf{s} = \mathbf{p}, \text{ if } \hat{c} \geq \tau, \quad \mathbf{s} = (1 - \alpha) \cdot \mathbf{p} + \alpha \cdot \mathbf{v}, \text{ if } \hat{c} < \tau, \quad (4)$$

and the final predicted stage is

$$\hat{y} = \mathrm{argmax}_j \, s_j \quad (5)$$

Here τ is a confidence threshold and $\alpha \in [0, 1]$ controls the influence of retrieved cases; we use $\tau = 0.7$ and $\alpha = 0.4$. When the federated model is confident ($\hat{c} \geq \tau$), its prediction is retained unchanged, and the retrieved cases act purely as supporting explanation. When the model is uncertain ($\hat{c} < \tau$), the regime in which classification errors are concentrated, the labels of clinically similar historical cases are allowed to re-rank the decision, so that strong agreement among close neighbours can correct an unreliable prediction.

This selective fusion is the reason the hybrid framework improves upon federated averaging without CBR: the CBR layer leaves confident predictions untouched and intervenes only on borderline cases, where neighbour evidence is most informative. The effect is therefore most pronounced under high-shift and domain-shift conditions (Table VI), where distributional differences lower model confidence and increase the proportion of cases routed through neighbour-based re-ranking.

Finally, to ensure the process is transparent to clinicians, the system exposes each step: for every new case, it presents the predicted stage together with the top-*k* retrieved cases and their confirmed diagnoses as supporting evidence. This step-by-step presentation allows clinicians to see not only the predicted stage but the historical cases underlying it, supporting validation and building trust in the system.

IV. Experimental Setup

A. Dataset Description

The proposed framework is evaluated using the publicly available LIDC-IDRI (Lung Image Database Consortium and Image Database Resource Initiative) dataset, which comprises approximately 1,018 thoracic CT scans collected from 1,010 patients with 2,669 annotated lung nodules provided by multiple expert radiologists. The dataset contains a diverse range of nodules with variations in size, shape, texture, and malignancy levels. In this work, the malignancy annotations are mapped to lung cancer stages (Stage I–IV) to facilitate multi-class stage prediction. To simulate a realistic multi-institutional healthcare environment, the dataset is partitioned into multiple subsets representing different clinical centers within the federated learning framework.

TABLE II. Simulated Federated Learning Environment(5 Hospitals Data)

Institution	Scans	Stage Distribution (I/II/III/IV)	Data Heterogeneity
Hosp1	200	40/30/20/10	Low variance
Hosp2	250	30/40/20/10	Medium (noise)
Hosp3	300	20/20/30/30	High (shift)
Hosp4	150	35/35/15/15	Domain shift
Hosp5	118	25/25/25/25	Balanced

Table II presents the simulated federated distribution of CT scan data across multiple participating institutions with varying stage distributions and heterogeneity conditions. Each simulated

institution performs local model training on its respective subset without sharing raw patient data, thereby preserving privacy while enabling collaborative learning. A non-IID data distribution is further simulated through label shifts across institutions, where certain centers contain a higher proportion of early-stage cases while others contain more advanced-stage samples. For each institution, the dataset is divided into 70% training, 15% validation, and 15% testing sets.

This experimental setup enables the evaluation of the proposed system under practical healthcare challenges such as distributed data ownership, institutional data heterogeneity, and privacy constraints. However, the stage labels are approximated using malignancy annotations and may not perfectly correspond to clinical staging standards. Additionally, due to limited access to real multi-institutional datasets, the federated learning environment is simulated through dataset partitioning.

B. Preprocessing and Feature Extraction

Before model training, the CT scan images go through a number of preparation procedures to ensure consistency, reduce noise-related variations, and improve feature quality. These preprocessing operations are essential for enhancing model robustness and diagnostic performance in medical imaging applications. Initially, lung nodule segmentation is performed using a U-Net architecture to isolate the lung regions and relevant nodules from surrounding anatomical structures[16]. Following segmentation, the extracted regions are resized to maintain uniformity across all CT scan samples (Table III).

TABLE III. SUMMARIZES THE COMPLETE EXPERIMENTAL CONFIGURATION AND HYPERPARAMETER SETTINGS USED FOR BOTH LOCAL TRAINING AND FEDERATED AGGREGATION

Category	Parameter	Value
Feature extractor	Backbone	3D ResNet-18
	Input size	64 × 64 × 64 (resized CT volume)
	Pretraining	Trained from scratch
Local training	Optimizer	Adam
	Learning rate	1×10^{-4}
	Batch size	16
	Local epochs per round	5
	Loss function	Categorical cross-entropy
Federated setup	Aggregation algorithm	FedAvg (weighted)
	Communication rounds	50
	Number of clients (hospitals)	5
	Client participation per round	Full (all clients)
	Data distribution	Non-IID (label shift across clients)
CBR module	Similarity metric	Cosine similarity
	Retrieved neighbours (k)	5
	Confidence threshold (τ)	0.7
	Fusion weight (α)	0.4
Data split	Train / Validation / Test	70% / 15% / 15%
Hardware	GPU	NVIDIA DGX B200

Next, Hounsfield Unit (HU) normalization is applied to reduce variations caused by differences in imaging equipment and acquisition settings. To further enhance image quality, basic noise reduction techniques, such as smoothing filters, are employed to suppress irrelevant artifacts while preserving important lung structures.

During training, data augmentation methods including rotation, flipping, and intensity shifting are used to increase model robustness and generalization. Since the proposed framework follows a federated learning paradigm, no central data pooling is performed, thereby preserving patient privacy across participating institutions.

After preprocessing, feature extraction is carried out using a 3D ResNet-18 architecture fine-tuned for multi-class lung cancer stage prediction. The network automatically learns discriminative imaging features such as nodule shape, texture, and intensity patterns from CT scans. These extracted deep features are subsequently utilized for both classification and case-based reasoning within the suggested framework. The proposed framework was implemented and evaluated using NVIDIA DGX B200 GPU-enabled computational infrastructure to support efficient deep learning training and federated learning experimentation on CT scan datasets.

C. Baseline Models

To assess the effectiveness of the proposed hybrid framework, it is compared against multiple baseline models representing different learning and deployment strategies.

1) Centralized deep learning model: In this baseline, all CT scan data is assumed to be available at a single location, and a CNN is trained in a centralized manner. This model serves as an upper-bound reference in terms of data availability, as it utilizes the complete dataset for training. However, it does not address privacy concerns associated with sharing sensitive medical data.

2) Federated learning without CBR: This baseline implements the federated learning framework using the same data distribution and training setup as the proposed method, but without incorporating the case-based reasoning module. It focuses solely on prediction performance while preserving data privacy[17]. This comparison helps isolate and evaluate the contribution of the CBR module in terms of interpretability and decision support.

3) Traditional machine learning models: Conventional machine learning algorithms, such as Support Vector Machines (SVM) and Random Forests, are used as additional baselines. These models are trained on extracted features from CT scans and provide a comparison with deep learning-based approaches [18]. Although less complex, they serve as a useful reference to highlight the advantages of the proposed hybrid framework.

To ensure a fair comparison, all deep-learning models, the proposed framework, FedAvg without CBR, the centralized CNN, and the standalone CNN — share the same 3D ResNet-18 backbone and are trained and evaluated on identical data splits with matched optimizer, learning rate, batch size, and training epochs.

The traditional machine-learning baselines (SVM and Random Forest) use the same 3D ResNet-18 features. This ensures performance differences reflect the methods rather than differing experimental conditions.

D. Evaluation Metrics

The suggested system's performance is assessed using both quantitative classification metrics and qualitative explainability measures to ensure a comprehensive assessment.

For classification performance, standard metrics such as accuracy, precision, recall, and F1-score are used. Since the proposed approach performs multi-class classification across different lung cancer stages, macro-averaged precision, recall, and F1-score are additionally considered to provide a balanced evaluation across all classes[15].

- Accuracy assesses the comprehensive accuracy of the model:

$$\frac{TP + TN}{TP + TN + FP + FN}$$

- Precision denotes the ratio of accurately predicted positive instances:

$$\frac{TP}{TP + FP}$$

- Recall (Sensitivity) evaluates the model's accuracy in identifying positive situations:

$$\frac{TP}{TP + FN}$$

- F1-score represents the harmonic mean of precision and recall:

$$2\,\frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

where, *TP*, *TN*, *FP,* and *FN* denote true positives, true negatives, false positives, and false negatives, respectively.

E. Qualitative Explainability Assessment

In addition to classification performance, the suggested system is assessed according to its capacity to provide meaningful and interpretable explanations using the case-based reasoning module.

1) Case relevance: Measures how closely the retrieved cases match the input case in terms of clinical and imaging features.

2) Interpretability: Assesses how easily clinicians can understand the reasoning behind the model's prediction based on the presented similar cases.

These qualitative aspects are important for determining the practical usability of the system in clinical settings, where transparency and trust play a crucial role in decision-making.

V. Results and Discussion

A. Qualitative Performance Analysis

The proposed framework performs stage-aware classification of lung cancer from CT scans across Stage I–IV. This section presents a qualitative and quantitative analysis of its performance, beginning with a comparison against baseline models and followed by a stage-wise breakdown and an evaluation of the case-based explanations.

1) Comparison with baselines: As shown in Table IV, the proposed Hybrid CBR-FL framework achieves the highest scores across all metrics (Macro F1 of 0.915 ± 0.002), with low standard deviations indicating stable performance across the four runs. It significantly outperforms all baselines ($p < 0.05$), including the centralized CNN despite the latter's access to pooled data and the federated baseline without CBR, confirming both the benefit of the hybrid design and the specific contribution of the case-based reasoning module.

Table V presents a simplified, imaging-based risk stratification scheme for NSCLC, included as clinical background rather than as a system output, to contextualize the stage-level predictions and their case-based explanations.

TABLE IV. Quantitative Comparison of Proposed and Baseline Models over Four Independent Runs (Mean ± Standard Deviation)

Model	Macro Precision	Macro Recall	Macro F1	p-value (F1 vs. Proposed)	Statistically Significant (p<0.05)?
Hybrid CBR-FL (Proposed)	**0.918 ± 0.003**	**0.912 ± 0.003**	**0.915 ± 0.002**	—	—
Centralized CNN	0.896 ± 0.006	0.892 ± 0.005	0.895 ± 0.005	0.0017	Yes
FedAvg (No CBR)	0.878 ± 0.006	0.873 ± 0.006	0.875 ± 0.005	0.0002	Yes
Standalone CNN	0.846 ± 0.004	0.841 ± 0.004	0.844 ± 0.003	< 0.0001	Yes
Local Avg	0.822 ± 0.007	0.814 ± 0.007	0.815 ± 0.007	< 0.0001	Yes

p-values are computed from an unpaired two-sample t-test of Macro F1 against the proposed model. All baseline differences from the proposed model are statistically significant ($p < 0.05$).

TABLE V. Simplified AI-Based Risk Stratification for NSCLC

Level	Risk	Benign Probability	Key Imaging Features	Growth Rate	Clinical Interpretation
Level A	Low	> 85% benign	Small, smooth, homogeneous nodules	Stable	Routine monitoring
Level B	Suspicious	40 – 85% benign	Medium size, mild irregularity, mixed texture	Slow growth	Follow-up/ possible biopsy
Level C	High	< 40% benign	Large, irregular, spiculated, heterogeneous	Rapid Growth	Immediate intervention

TABLE VI. STAGE-WISE BREAKDOWN (PROPOSED MODEL)

Stage	Precision	Recall	F1-Score
Stage I	0.95	0.94	0.945
Stage II	0.93	0.92	0.925
Stage III	0.89	0.88	0.885
Stage IV	0.9	0.9	0.9

Table VI presents the stage-wise performance breakdown of the proposed Hybrid CBR-FL framework across different lung cancer stages. The outcomes show that the model maintains balanced precision, recall, and F1-score values for Stage I–IV classification, indicating reliable and consistent stage-aware prediction capability. Fig. 3 presents the comparative macro-level performance analysis of different classification models evaluated on the LIDC-IDRI federated simulation dataset. The proposed Hybrid CBR-FL framework achieves the highest overall performance, demonstrating improved classification effectiveness and robustness compared to conventional federated and standalone approaches.

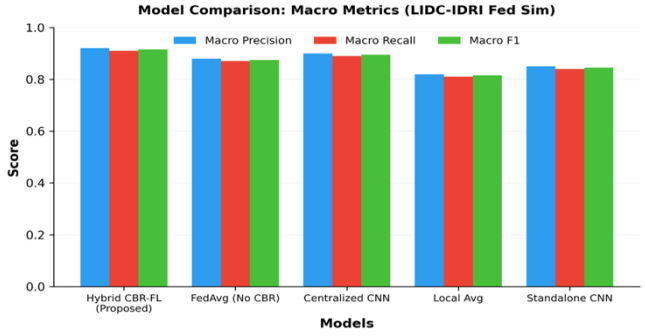

Fig. 3. Macro performance comparison across models.

B. Comparative Performance Analysis

Table VII presents the comparative performance analysis of the proposed Hybrid CBR-FL framework and the conventional FedAvg approach under different institutional heterogeneity conditions. The findings show that the suggested framework regularly achieves greater F1-scores across all simulated hospital environments. The performance improvement is particularly significant in high-shift and domain-shift scenarios, where variations in data distribution and imaging characteristics are more prominent. This indicates that the proposed hybrid framework is more robust and adaptable to heterogeneous decentralized healthcare data compared to traditional federated learning approaches.

TABLE VII. PERFORMANCE COMPARISON UNDER DIFFERENT FEDERATED DATA HETEROGENEITY CONDITIONS

Institution Heterogeneity	Hybrid CBR - FL F1	FedAvg F1	Improvement
Low Variance (Hosp1)	93.2%	91.8%	+1.4%
High Shift (Hosp3)	89.1%	82.3%	+6.8%
Domain Shift (Hosp4)	87.6%	80.5%	+7.1%
Macro Average	91.4%	87.5%	+3.9%

C. Explainability Evaluation

The final output of the proposed framework consists of the predicted lung cancer stage (Stage I–IV), which is further supported by retrieval of clinically similar historical cases through the Case-Based Reasoning module. The CBR module also improves the interpretability of lung cancer predictions. Instead of providing only a final classification result, the system retrieves previously similar lung nodule cases from the dataset based on extracted imaging features. The explainability component assists in highlighting similarities between the current CT scan and historical cases with related malignancy characteristics and cancer stages[19]. By presenting comparable cases, the framework supports clinician-oriented decision-making and improves trust in the automated diagnostic process. The retrieved similar cases help provide supportive evidence for the predicted stage classification while maintaining patient data privacy across decentralized institutions.

TABLE VIII. QUANTITATIVE EVALUATION OF THE CASE-BASED REASONING MODULE (COMPUTED OVER THE FULL TEST SET, N = 75)

Metric	Value
Case Relevance (mean top-k cosine similarity)	0.842
Stage-match rate of retrieved cases (%)	76.4
Retrieval precision@k	0.764

TABLE IX. EXAMPLE OF CASE-BASED EXPLANATIONS PRODUCED BY THE PROPOSED FRAMEWORK FOR THREE REPRESENTATIVE TEST CASES

Case	Predicted Stage	Top-k Retrieved Cases (Stage, Similarity)	Mean Similarity	Stage-Match
Patient 1	Stage II	(II, 0.92), (II, 0.90), (II, 0.88), (III, 0.85), (I, 0.80)	0.870	3/5 (0.60)
Patient 2	Stage I	(I, 0.95), (I, 0.94), (I, 0.91), (I, 0.89), (II, 0.86)	0.910	4/5 (0.80)
Patient 3	Stage III	(III, 0.88), (III, 0.85), (III, 0.83), (III, 0.80), (III, 0.78)	0.828	5/5 (1.00)

TABLE X. COMPARISON OF EXPLANATION METHODS FOR THE FEDERATED MODEL

Method	Explanation type	What the clinician sees	Localization
Grad-CAM	Saliency heatmap	Highlighted image regions	Diffuse
CBR (Proposed)	**Retrieved similar cases**	**Confirmed prior cases + stages**	**Indirect**

D. Quantitative Explainability Evaluation

Beyond qualitative description, the Case-Based Reasoning module is evaluated using two measures: Case Relevance, the mean cosine similarity of the top-k retrieved cases to the query, and the stage-match rate, the fraction of retrieved cases whose confirmed stage matches the predicted stage. Over the full test set (N = 75), the framework achieves a Case Relevance of 0.842 and a stage-match rate of 76.4% (Table X), showing that the retrieved cases are both visually similar to the query and, in most instances, clinically consistent with its predicted stage. Table X further illustrates this on three representative cases: for Patient 1

(predicted Stage II), three of five retrieved cases confirm Stage II, while for Patient 3 (predicted Stage III), all five are Stage III with high similarity. In each case, the clinician sees not only a predicted stage but the specific prior cases and confirmed diagnoses supporting it, enabling direct inspection and validation of the reasoning.

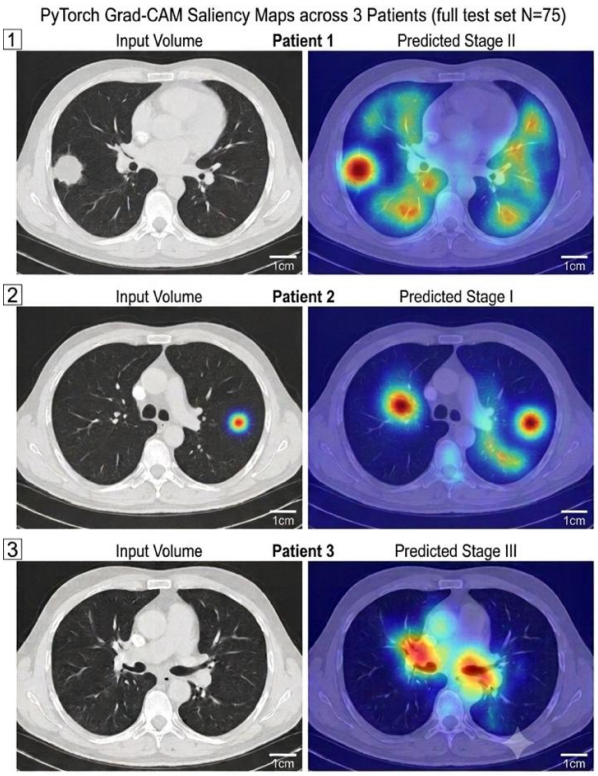

Fig. 4. Grad-CAM saliency maps for three test cases (Patients 1–3; predicted Stages II, I, III), each CT slice shown beside its Grad-CAM activation from the 3D ResNet-18. Activations are diffuse, in contrast to the case-level evidence from the CBR module.

Comparison with gradient-based explanation. To benchmark the proposed explanations against an established technique, Grad-CAM was applied to the same 3D ResNet-18 model at its final convolutional block. The resulting saliency maps Fig. 4. are often diffuse and multifocal, spanning several regions across both lungs rather than isolating a single decisive finding. Such maps indicate where the model attended but not why a stage was assigned, and are difficult to translate into a clinical rationale. In contrast, the CBR module grounds each decision in specific, previously confirmed cases with known stages (Tables VIII and IX), offering concrete, inspectable evidence. As summarized in Table X, the two approaches are complementary; Grad-CAM provides spatial attribution, while CBR provides justification grounded in prior confirmed cases, which is well suited to stage-aware diagnosis.

E. Discussion

The proposed hybrid framework demonstrates significant potential for supporting intelligent and privacy-aware lung cancer diagnosis in clinical environments. By integrating deep learning, federated learning, and case-based reasoning, the system is capable of assisting clinicians in performing stage-aware classification using CT scan images. The inclusion of explainability through similar case retrieval further improves the transparency of the diagnostic process, which is especially crucial in healthcare applications where interpretability and trust are essential for decision-making. This explainability is supported quantitatively, with retrieved cases showing high similarity and stage consistency, and compares favourably with gradient-based saliency methods in providing clinically inspectable evidence.

1) Privacy benefits: One of the main benefits of the suggested strategy is its privacy-preserving capability. The federated learning framework permits cooperative model training across several institutions without exchanging raw patient data, in contrast to traditional centralized learning techniques. This reduces privacy risks and supports secure utilization of distributed healthcare datasets while maintaining data confidentiality and institutional ownership.

2) Deployment feasibility: From a deployment perspective, the framework is designed to operate in decentralized healthcare environments where data availability and imaging characteristics may vary across institutions. The ability of the system to learn from heterogeneous distributed datasets improves its adaptability and robustness in practical medical settings. Although the current work utilizes simulated federated institutions and approximated stage annotations, the proposed architecture provides a strong foundation for future integration with real-world clinical infrastructures and multi-institutional healthcare systems.

VI. Conclusion and Future Work

In this work, we presented a hybrid framework that brings together Case-Based Reasoning and Federated Learning for stage-aware lung cancer classification from CT images. The design tackles two problems that usually pull against each other: it keeps patient data private by training across institutions without ever moving raw scans, while still holding up its classification performance even when the data across those institutions is uneven.

Across our experiments, the proposed model consistently outperformed local learning, standalone CNNs, and conventional FedAvg on macro-averaged precision, recall, and F1-score, and it did so while classifying cases across Stage I–IV and staying stable under non-IID, multi-institutional data. Just as importantly, the case-based reasoning component earned its place beyond raw accuracy: the cases it retrieves are genuinely similar to the query and largely consistent in stage, giving clinicians concrete, inspectable evidence for a prediction — something the diffuse heatmaps of gradient-based methods struggle to offer.

Taken together, these results suggest that pairing federated learning with case-based reasoning is a practical route toward lung cancer diagnosis that is private, accurate, and understandable at the same time — and a promising foundation for the kind of clinical decision-support systems that hospitals could actually trust and adopt.

A. Limitation

Despite the promising performance of the proposed framework, certain limitations remain. The federated learning environment considered in this study is simulated through dataset partitioning and does not fully represent real-world multi-institutional clinical infrastructures. In addition, the lung cancer stage labels are approximated using malignancy annotations available in the LIDC-IDRI dataset, which may not completely correspond to standard clinical staging procedures. A further constraint is the scale of the data: the 2,669 annotated nodules distributed across five simulated institutions and four stage categories are modest for deep federated training, which increases the risk of overfitting and limits the statistical power of per-stage estimates. Data augmentation (rotation, flipping, and intensity shifting) was applied to partially mitigate this**.** Furthermore, the current evaluation is limited to retrospective publicly available CT scan datasets, and external clinical validation with larger heterogeneous patient populations is necessary to further assess the generalizability and practical applicability of the proposed system in real healthcare settings.

B. Future Work

Future work will extend the proposed framework toward real-world hospital deployment to evaluate its performance under practical clinical and regulatory conditions. In addition, a radiologist-in-the-loop feedback mechanism will be incorporated to improve case retrieval quality, explanation reliability, and model refinement. Finally, the framework will be expanded to support multi-modal data integration, combining CT imaging with clinical and pathological information to enhance robustness and diagnostic accuracy.

REFERENCES

[1] Yu, M., Tyson, C., Limburg, P. J., & Beer, T. M. (2026). A quantitative framework to assess the potential of earlier cancer detection to improve cancer survival. *Cancer Epidemiology, Biomarkers & Prevention*, OF1–OF6.

[2] Lazzari, C., Spagnolo, C. C., Ciappina, G., Di Pietro, M., Squeri, A., Passalacqua, M. I., Marchesi, S., Gregorc, V., & Santarpia, M. (2023). Immunotherapy in early-stage non-small cell lung cancer (NSCLC): Current evidence and perspectives. *Current Oncology*, *30*(4), 3684–3696.

[3] Gayap, H. T., & Akhloufi, M. A. (2024). Deep machine learning for medical diagnosis, application to lung cancer detection: A review. *BioMedInformatics*, *4*(1), 236–284.

[4] Thanoon, M. A., Zulkifley, M. A., Mohd Zainuri, M. A. A., & Abdani, S. R. (2023). A review of deep learning techniques for lung cancer screening and diagnosis based on CT images. *Diagnostics*, *13*(16), 2617.

[5] Sari, S., Sutikno, T., Soesanti, I., & Setiawan, N. A. (2023). A review of convolutional neural network-based computer-aided lung nodule detection system. *IAES International Journal of Artificial Intelligence*, *12*(3), 1044–1061.

[6] Saha, A., Ganie, S. M., Pramanik, P. K. D., Yadav, R. K., Mallik, S., & Zhao, Z. (2024). VER-Net: A hybrid transfer learning model for lung cancer detection using CT scan images. *BMC Medical Imaging*, *24*(1), 120.

[7] AR, N. K. (2025). Proximal guided hybrid federated learning approach with parameter efficient adaptive intelligence for pneumonia diagnosis. *Scientific Reports*, *16*(1), 2412.

[8] Saini, S. S. (2025). Federated learning for privacy-preserving healthcare AI models. *International Journal of Unified Research & Development (IJURD)*, *1*(1), 1–5.

[9] Oyelade, O. N., & Ezugwu, A. E. (2020). A case-based reasoning framework for early detection and diagnosis of novel coronavirus. *Informatics in Medicine Unlocked*, *20*, 100395.

[10] Ali, S. I., Jung, S. W., Bilal, H. S. M., Lee, S. H., Hussain, J., Afzal, M., Han, M., Kim, T. S., Hwang, K. Y., & Lee, S. (2022). Clinical decision support system based on hybrid knowledge modeling: A case study of chronic kidney disease-mineral and bone disorder treatment. *International Journal of Environmental Research and Public Health*, *19*(1), 226.

[11] Mustafa, E. M., Saad, M. M., & Rizkallah, L. W. (2023). Building an enhanced case-based reasoning and rule-based systems for medical diagnosis. *Journal of Engineering and Applied Science*, *70*(1), 139.

[12] Tofighi, A. B., Ahmadi, A., & Mosadegh, H. (2025). A novel case-based reasoning system for explainable lung cancer diagnosis. *Computers in Biology and Medicine*, *185*, 109547.

[13] Pradeep, P., Caro-Martínez, M., & Wijekoon, A. (2024). A practical exploration of the convergence of Case-Based Reasoning and Explainable Artificial Intelligence. *Expert Systems with Applications*, *255*, 124733.

[14] Liu, Y., Huang, J., Chen, J.-C., Chen, W., Pan, Y., & Qiu, J. (2024). Predicting treatment response in multicenter non-small cell lung cancer patients based on federated learning. *BMC Cancer*, *24*, 688.

[15] Heidari, A., Javaheri, D., Toumaj, S., Navimipour, N. J., Rezaei, M., & Unal, M. (2023). A new lung cancer detection method based on the chest CT images using federated learning and blockchain systems. *Artificial Intelligence in Medicine*, *141*, 102572.

[16] Turjya, S. M., & Fawakherji, M. (2026). Federated lung nodule segmentation using a hybrid Transformer–U-Net architecture. *Scientific Reports*, *16*(1), 5228.

[17] Chen, Y., Liu, L., Feng, B., Chen, Y., Xu, J., & Chen, X. (2025). A meta-learning-based robust federated learning for diagnosing lung adenocarcinoma and tuberculosis granulomas. *Frontiers in Oncology*, *15*, 1666937.

[18] Causey, J. L., Guan, Y., Dong, W., Walker, K., Qualls, J. A., Prior, F., & Huang, X. (2019). *Lung cancer screening with low-dose CT scans using a deep learning approach*. arXiv preprint arXiv:1906.00240v

[19] Wang, J., Gao, R., Huo, Y., Bao, S., Xiong, Y., Antic, S. L., Osterman, T. J., Massion, P. P., & Landman, B. A. (2019). *Lung cancer detection using co-learning from chest CT images and clinical demographics*. arXiv. preprint arXiv:1902.08236.