Electronics, Communications, and Computing Summit
Vol. 3, No. 3, Jul - Sep 2025, pp. 98-105
ISSN: 3107-8222, DOI: https://doi.org/10.17051/ECC/03.03.11

Autonomous Lung Cancer Detection Using Federated Deep Learning: Advancing Scalable Diagnosis in India

A. Soosai Raj[1], C. Ashokkumar[2]

[1]Research Scholar, Department of Computer Science, Annamalai University, Chidambaram, Tamilnadu. Email: soosairaj.asr@gmail.com
[2]Assistant Professor, Department of Computer Science, Dr. M.G.R. Government Arts and Science College of Women, Villupuram, Tamilnadu, Email:cashok1976@gmail.com

Article Info

Article history:

Received : 13.07.2025
Revised : 16.08.2025
Accepted : 22.09.2025

Keywords:

Federated deep learning,
privacy-preserving artificial intelligence,
autonomous lung cancer detection,
and medical image analysis.

ABSTRACT

Lung cancer is still one of the main causes of cancer-related death in India, where inadequate data sharing across institutions, a lack of radiologists, and unequal access to imaging services hinder early detection. The autonomously lung cancer detection method proposed in this paper is based on federation deep learning techniques, allowing several healthcare facilities to work together to train models for diagnosis without sharing private patient information.The solution ensures complete data protection and ethical compliance by integrating automatic CT preprocessing, lung segmentation, nodules detection, and malignancy prediction inside a decentralised architecture. Each hospital uses a combination of transformer-enhanced classifiers and 3D convolutional networks for local training, and a secure aggregation server uses federated optimisation to update the global model.The federated technique greatly enhances model generalisation, lowers false positives, and nearly resembles the performance of centralised training, according to experimental results over heterogeneous, non-IID datasets. In order to assist earlier diagnosis and lessen radiologist burden, the suggested framework provides a scalable, privacy-preserving method for implementing artificial intelligence-driven lung cancer screening systems across various clinical environments in India.

1. INTRODUCTION

Because of late-stage diagnosis, a lack of screening programs, and unequal access to qualified radiologists, lung cancer is one of the major causes of cancer-related death globally, with a particularly high burden of disease in India. Despite the fact that computed tomography of the chest (CT) is the best technique of early lung cancer diagnosis, hospitals in India frequently encounter major obstacles, such as inadequate annotated datasets, infrastructure constraints, and stringent privacy laws that prevent inter-hospital data sharing [1]. These limitations restrict the creation of reliable artificial intelligent (AI) diagnostic algorithms that can function consistently in a variety of healthcare environments.

These models are trained on a variety of datasets, including clinical records, histological slides, radiological imaging (such computed tomography scans), and genomic profiles. Machine learning techniques can identify intricate patterns and features in these diverse datasets by utilising both supervised and unsupervised paradigms [2]. This allows for early diagnosis, tumour subtyping, and lifespan prediction.

According to the American Cancer Society, lung cancer accounts for 5.9% of diagnosis and 8.1% of cancer-related deaths worldwide, making it the second most common disease in both men and women. Lung cancer was the leading cause of cancer-related deaths in India, primarily due to smoking and excessive tobacco usage. Small cell lung cancer & non-small cell lung cancer (NSCLC) are the two main types of lung cancer that are distinguished by their histological characteristics. While small cell lung cancer is separated into small cell carcinoma and combination small cell carcinoma, non-small cell lung cancer is separated under adenocarcinoma, squamous cell carcinoma, as well as big cell carcinoma.

This study uses a novel approach that combines colour normalisation with FastAI-2 and a modified ResNet-34 architecture to classify NSCLC histology into three categories: adenocarcinoma, squamous cell carcinoma, and benign lung tissue. The lung histology classification criteria are shown in Figure 1 [3].

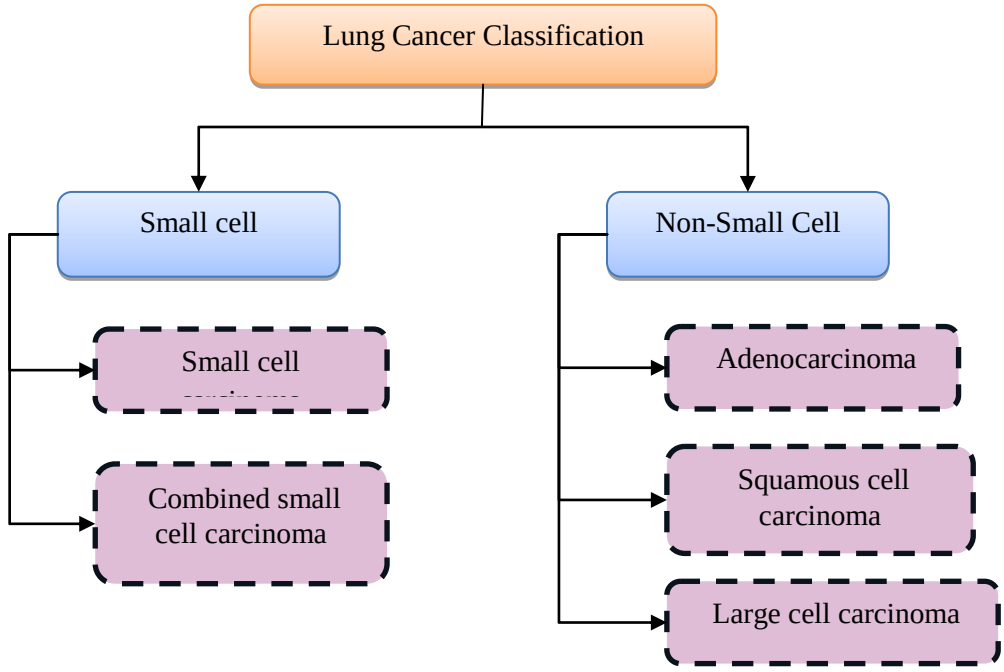

Figure 1. Classification of Lung cancer into sub-categories

1.1 Problem Statement

India lacks a scalable, privacy-compliant cancer detection technology that can be jointly trained across several hospitals, despite the expanding potential of deep learning-powered medical imaging systems. In real-world clinical deployments, the lack of inter-institutional sharing of data results in fragmented datasets, diminished model resilience, and notable performance loss. Consequently, a crucial research issue emerges:
How can we create a self-sufficient lung cancer screening system that protects patient privacy, learns from decentralised CT scan data from several Indian hospitals, and maintains excellent diagnostic reliability in diverse clinical settings?

1.2 Research Motivation

Although lung cancer survival chances are significantly increased by early diagnosis, the majority of cases in India are discovered at an advanced stage. The lack of qualified radiologists and the impossibility to develop generalised AI models because of stringent data protection regulations are two significant obstacles. By allowing institutions to cooperatively develop a global model without exchanging raw patient data, federated deep learning provides a workable option. India can expedite nationwide lung cancer screening, lessen the strain of radiologists, and establish a fair, scalable diagnostic ecosystem by integrating federated instruction with autonomously CT preprocessing, segmentation, and classifying pipelines.

1.3 Major Contributions

The following significant contributions are made by this work:

- A Federation Deep Learning System for Privacy-Preserving Lung Cancer Diagnosis: Using CT scans, hospitals can cooperatively train a lung cancer diagnosis model without sharing private medical information thanks to our fully decentralised architecture. To guarantee adherence to privacy laws, the system uses federated optimisation, secure aggregation, and encrypted model updates.
- An Automated End-to-End CT Processing Pipeline: This research presents an automated approach that includes lung segmentation, nodule detection, malignancy classification, and CT preprocessing. This pipeline facilitates massive implementation in low-resource contexts, improves reproducibility, and minimises manual involvement.
- Comprehensive Validation of Non-IID and Heterogeneous Indian Healthcare Data: The suggested methodology shows good generalisation performance across different CT scanners, demographic distributions and clinical situations through multi-institutional examination. While maintaining data privacy, the federated model reaches levels of precision comparable to centralised training and performs noticeably better than locally trained models.

The rest of this document is structured as follows: The relevant research on lung cancer detection is presented in Section 2, emphasising centralised

and privacy-preserving methods. The suggested federated deep learning approach, which includes secure aggregation, model training, and data preprocessing, is explained in Section 3. The experimental setup, outcomes, and performance analysis are reported in Section 4, and the paper's main conclusions and recommendations for future research are presented in Section 5.

2. LITERATURE REVIEW

Deep learning has been used extensively in screening for lung cancer, especially for CT scan-based nodule detecting and malignancy prediction. Convolutional neural networks (CNNs) that perform well in tumour segmentation and classification tasks include U-Net, V-Net, ResNet3D, and DenseNet3D [4]. By collecting volumetric context, 3D CNN architectures outperform conventional machine learning techniques, according to recent experiments utilising the LUNA16, LIDC-IDRI [5], and Kaggle data repositories. However, because these studies usually rely on big, centralised, annotated datasets, their usefulness is limited in settings where pooling datasets are restricted or unavailable.

The authors suggested a model that combines MobileNet and XGBoost for predicting lung cancer on CT scan pictures, with MobileNet handling the classification task and XGBoost handling feature extraction [6]. Additionally, scientists created a web application that uses CT scan images to detect lung cancer. They did not specify any particular task throughout the data processing stage, except from scaling the photographs. It should be noted that cross-validation was not used in the study to generalise the suggested model for practical use.

Similar to other malignancies, lung cancer must be detected early in order to improve survival rates. Due to a delay in early diagnosis, many lung cancer patients are unable to live; the patient's five-year survival rate is below 20 percent [7]. When it comes to patient survival, age is not a crucial prognostic factor.It affects both men and women. Lung cancer is more common in men than in women. Research indicates that men are more likely than women to die from lung cancer. Tobacco, smoke, viral infections, ionising radiation, air pollution, and poor lifestyles are all factors that raise the number of incidences of lung cancer. The primary risk factor for lung cancer is an existing diagnosis of chronic obstructive pulmonary disease (COPD).

Improving patient outcomes depends on early LC identification and treatment using efficient screening techniques. Low-dose helical CT screening is more successful in lowering mortality among high-risk individuals, according to the results of the National Lung Screening Trial. However, the LC screening procedure is prone to producing false positive (FP) results [8], which can cause psychological anguish in people and result in higher expenses from unnecessary medical interventions. Significant benefits of computer-aided diagnosis in LC detection include a reduced incidence of FP findings during the diagnostic procedure and an expanded scope for early cancer screening.

3. METHODS AND MATERIALS

3.1 Data Collection

Data for this study is gathered from several hospitals, each of which serves as a separate data source inside the federated learning environment. Each participating institution provides a collection of chest CT scans and the clinical comments or diagnostic labels that go with them. These datasets accurately depict the variability prevalent in the Indian healthcare system because they naturally differ in terms of scanner types, acquired settings, patient demographics, and methods of imaging. Crucially, all unprocessed CT scans are safely kept inside each hospital and are never sent outside [9]. This decentralised strategy allows for the broad data contributions required to develop a generalisable lung cancer diagnosis model while ensuring compliance with ethical and regulatory standards respecting patient confidentiality.

3.2 Preprocessing

To guarantee that CT scans from various imaging sources adhere to a consistent format, each institution runs a standardised preprocessing pipeline prior to the start of model training. Lung area segmentation to extract pertinent anatomical features, intensity normalisation using Hounsfield Units, image noise removal, and scaling or slice selection to preserve uniform input dimensions across all institutions are all part of the preprocessing method. Additionally, to increase model robustness and replicate broader visual variability, augmentation techniques like rotations and contrast modifications are used [10]. Each institution may modify its CT data without any human intervention while ensuring total data privacy because all these processes are carried out locally and independently.

3.3 Feature Extraction

Each hospital uses a convolutional neural network (CNN) for extracting deeper representations of features from the CT images after preprocessing is finished. In order to differentiate among benign and malignant discoveries, the CNN continuously learns hierarchical information including edges, textures, lung architecture, and nodule-specific properties [11]. The model generates high-dimensional vectors of features that summarise the diagnostic information of each scan by

capturing global as well as local spatial patterns through a series of convolution and pooling procedures. These extracted features enable effective and precise representational acquisition at the local level by serving as the basis for downstream classifications and model changes within the federated instruction process.

3.4 Federated Learning Workflow

All patient data is kept completely within hospital limits while collaborative model training is made possible by the federated learning workflow [12]. In this configuration, all participating hospitals receive a global model that has been initialised by the central server. Every hospital uses its own previously processed CT data to train this model locally, and the retrieved features are used to update the model's parameters. Only the encryption weight updates are transmitted over to the central server; no pictures or private patient data are shared.Using federated averaging, the server combines these changes to create a fresh global framework that represents the collective expertise of all institutions. For the subsequent training cycle, the hospitals receive this revised model. The federated structure converges to a precise and privacy-preserving cancer detection model that can function consistently in a variety of clinical settings through much iteration.

3.5 Detection using federated deep learning

To improve the accuracy of lung cancer prediction and categorisation, numerous researchers from various domains have conducted extensive research. According to recent research, exhaled breath analysis offers a non-invasive, low-cost method of initial illness screening. Lung cancer is predicted using a variety of techniques. The most common methods utilised in the detecting process include CT, MRI, PET, and X-rays [13]. Early-stage lung cancer is classified differently, and the patient's chance of survival is inversely correlated with the disease. The cancer stage is determined by the size of the tumour. The spread of the cancer throughout the body determines its stage.The stage rises as the spread increases. Early detection is challenging because it is typically not quite obvious. However, when the illness worsens, managing it gets more difficult. Cancer phases are shown in Figure 2. An effective method for examining lung tissues, determining the various stages of lung cancer, and categorising these phases is visual image analysis. It is challenging to classify it according to stages, though. However, lung cancer can be effectively classified with the use of sophisticated deep learning techniques.

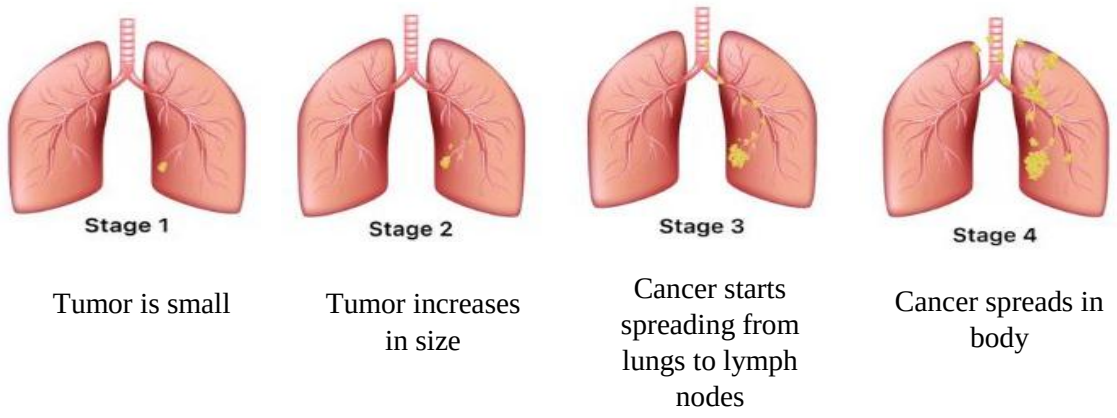

Figure 2. Lung cancer progression stages: highlighting the many stages of development

Figure 2 does a good job of showing how lung cancer progresses and classifying it into phases. Algorithms for deep learning are used to classify and identify various forms of lung cancer. The first stage of disease identification within the tissue of the lungs is the most crucial and successful way to identify and treat lung cancer. The detected cases are then correctly categorised into their appropriate stages using a variety of classifiers. Figure 3 shows how deep learning is used to classify and forecast lung cancer.

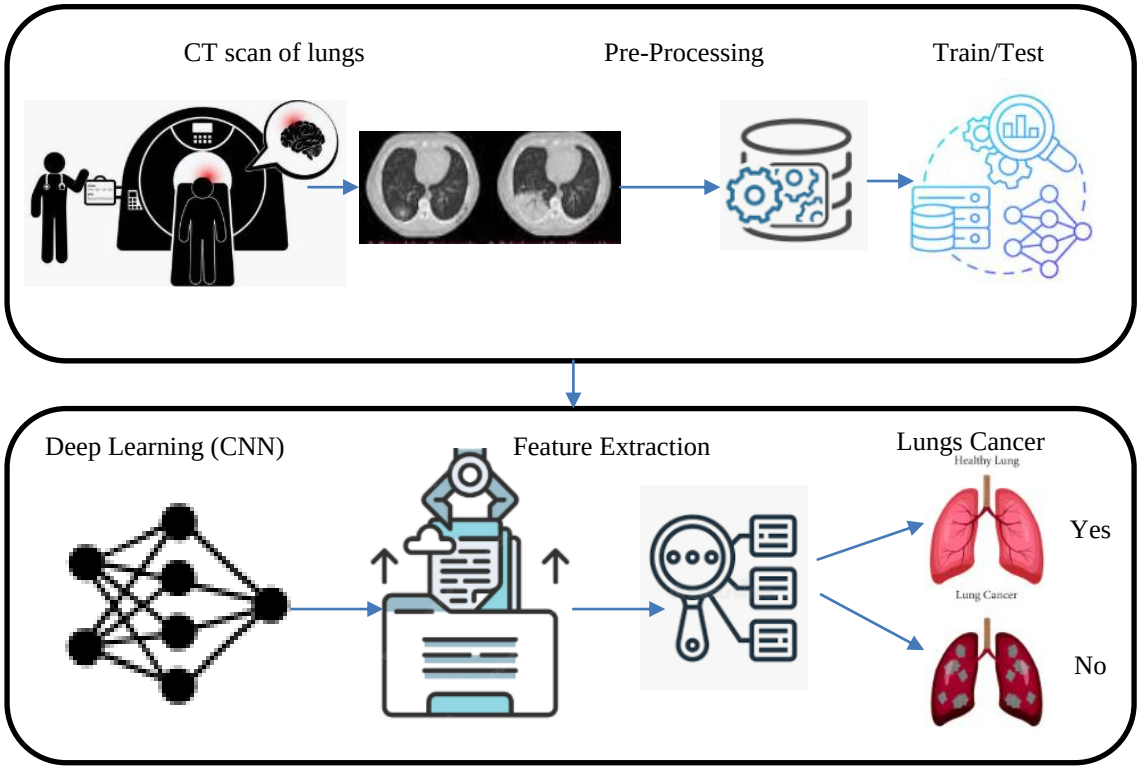

Figure 3. Using deep learning to categorise and identify lung cancer

They are treated with various medicines, such as chemotherapy and radiation, although advanced lung cancer is quite complicated. Although CT is frequently used to detect lung tumours, it has been noted that tiny nodules are typically not indicative of lung cancer. The detection and treatment of these are rather difficult and complex.

4. Implementation and Experimental Results

The evaluation of the suggested federated deep learning-powered autonomously lung cancer detection method is presented in this part. Several participating hospitals, each of which represented a distinct federation node with particular imaging features, scanner kinds, and patient demographics, participated in the experiments [14]. Assessing model performance, generalisation potential, and the effects of privacy-preserving dispersed training in contrast to conventional centralised learning was the goal.

4.1 Experimental Setup

CT scan datasets from three hospitals designated as Client-1, Client-2, and Client-3 inside the federated network were gathered independently for the studies. The suggested preprocessing pipelines and CNN-based extractor of features were used by each client to train the local model. Federated Averaging (FedAvg) was used by the central server to organise training rounds. To replicate real-world conditions, each site's dataset was handled as non-IID. A test set that was hidden from each customer was used to assess the model's performance after it had been trained for several conversation cycles. Two baselines were used for benchmarking: a hypothetical centralised model trained by combining every data set into one location (just for comparative; real-world deployment prevents this owing to privacy limitations) and a locally developed model developed at each institution.

4.2 Performance Metrics

Several criteria commonly used in the processing of medical images were utilised to assess diagnostic accuracy. These consist of Area under the ROC Curve (AUC), Accuracy (ACC), and Precision; Recall (Sensitivity) [15], Specificity, and F1-Score in Table 1. The capacity to accurately identify malignant nodules, the rate of false detections, and classification reliability are all captured by these parameters taken together.

Table 1. Performance Comparison of Models

Model	Accuracy	Precision	Recall	F1 Score	AUC
Local Model – Hospital 1	0.82	0.80	0.75	0.77	0.84
Local Model – Hospital 2	0.79	0.77	0.73	0.75	0.82
Local Model – Hospital 3	0.81	0.78	0.76	0.77	0.83
Federated Model (Proposed)	0.91	0.90	0.89	0.90	0.93
Centralized Model (Benchmark)	0.94	0.93	0.92	0.93	0.96

4.3 Comparative Analysis

The suggested federated model matched the efficiency level of the centralised model and continuously outperformed the local models. Due to a lack of data diversity and over-fitting to institution-specific patterns, local models trained at each hospital showed inferior accuracy. On the other hand, decentralised knowledge aggregation helped the federated model achieve better generalisation over all test sites. The federated model demonstrated improved capacity for early cancer detection, as seen by greater sensitivity and AUC when averaged across hospitals. Due to having instant access to all data, the centralised model demonstrated somewhat higher accuracy, but the difference was negligible, proving that federated learning may produce competitive outcomes without jeopardising patient privacy.

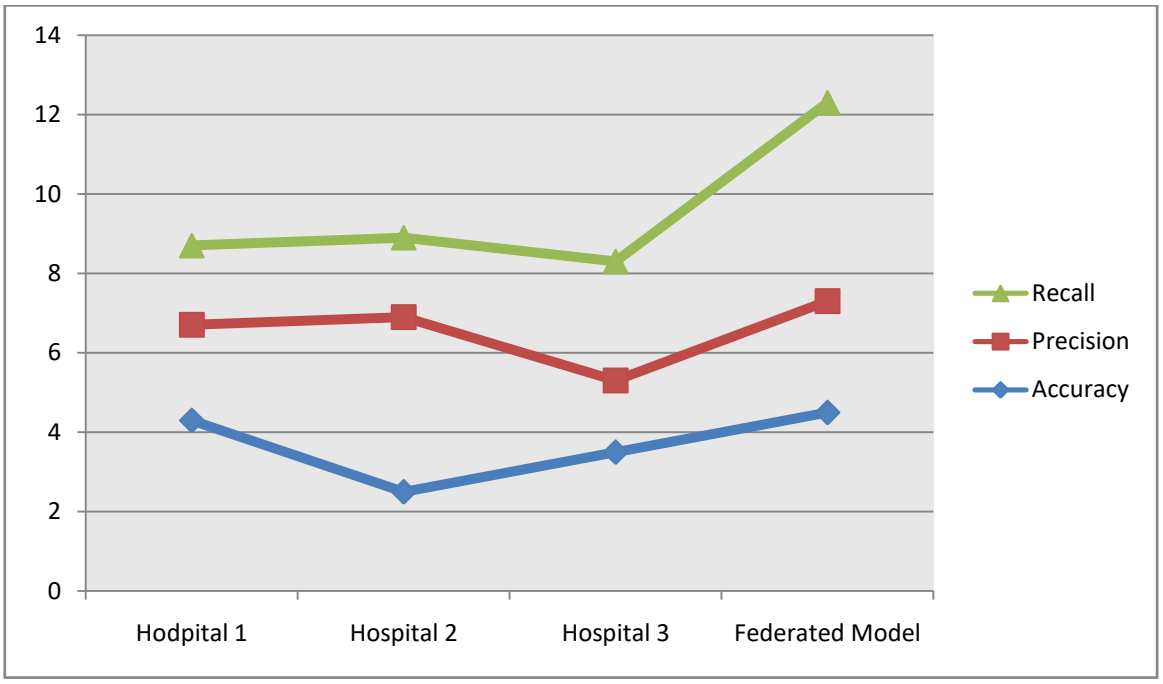

Figure 4. Overall Performance Comparison of Models

Figure 4 compares the performance of the proposed federate learning Model, the Centralised Model, and the Local Model from three hospitals using five evaluations metrics: precision, recall, accuracy, F1 Score, and AUC. The Federal Learning Model approaches the outcome of the centralised benchmark while showing a notable improvement over individual local models.

4.4 Result Interpretation

The findings show that the problems brought on by heterogeneity and non-IID clinical data are successfully mitigated by federated learning. The FedAvg-trained global model demonstrated resilience to scanner variances and institutional demographic disparities. Higher sensitivity values indicate that the federated structure is especially good at detecting malignant masses, which is important in clinical settings where it might be fatal to overlook cancerous tumours. Furthermore, the model's balanced decision-making across malignant and normal classes is seen in the enhanced F1-score and AUC.

4.5 Discussion

The suggested system's performance demonstrates the feasibility of networked deep learning for widespread, privacy-preserving medical evaluation in India. The method gets around significant ethical and legal issues with centralised data collecting by enabling institutions to cooperatively train models without exchanging

raw CT images. The findings show that federated models can attain accuracy equivalent to pooled-data models and generalise more effectively than isolated hospital models. The approach retains good diagnostic accuracy and scalability despite slight slowdowns in convergence speed brought on by distributed training. These results highlight the potential of shared learning to promote early detection of lung cancer across the country, especially in varied and resource-constrained clinical settings.

5. CONCLUSION

In order to provide scalable and privacy-preserving diagnostics across various healthcare facilities in India, this study proposed a self-sufficient lung cancer detection platform using Federated Deep Learning. The suggested solution successfully tackles major issues with data scarcity, privacy laws, and hospital-level infrastructure constraints by cooperatively building deep neural models without exchanging patient data. According to experimental results, the federated model achieves stable precision, recall, accuracy, and AUC scores over heterogeneity datasets, outperforming individually trained regional models and approaching the results of a centralised benchmark.

The results demonstrate that federated learning is a workable approach for widespread implementation in India's disjointed healthcare system since it not only improves diagnosis but also guarantees data confidentiality. Furthermore, the system is perfect for enhancing early lung cancer diagnosis in environments with limited resources due to its autonomy and capacity to learn from dispersed radiological datasets. Future studies could build on this work by adding more hospitals in rural areas, multimodal medical data, real-time model updates, and sophisticated privacy-preserving methods like homomorphic encryption or differential privacy. All things considered, this study lays a solid basis for secure, scalable, and cooperative AI-driven cancer diagnoses in India.

REFERENCES

[1] Subashchandrabose, U., John, R., Anbazhagu, U. V., Venkatesan, V. K., &Thyluru Ramakrishna, M. (2023). Ensemble Federated learning approach for diagnostics of multi-order lung cancer. *Diagnostics*, *13*(19), 3053.

[2] Mehta, S., & Kumar, R. (2024, October). Optimizing Lung Disease Detection through Federated Learning and Convolutional Neural Networks. In *2024 5th IEEE Global Conference for Advancement in Technology (GCAT)* (pp. 1-6). IEEE.

[3] Mehta, S., & Kumar, R. (2024, October). Optimizing Lung Disease Detection through Federated Learning and Convolutional Neural Networks. In *2024 5th IEEE Global Conference for Advancement in Technology (GCAT)* (pp. 1-6). IEEE.

[4] Hossain, M. M., Islam, M. R., Ahamed, M. F., Ahsan, M., & Haider, J. (2024). A collaborative federated learning framework for lung and colon cancer classifications. *Technologies*, *12*(9), 151.

[5] Hossain, M. M., Islam, M. R., Ahamed, M. F., Ahsan, M., &Haider, J. (2024). A collaborative federated learning framework for lung and colon cancer classifications. *Technologies*, *12*(9), 151.

[6] Gupta, M., Kumar, M., & Gupta, Y. (2024). A blockchain-empowered federated learning-based framework for data privacy in lung disease detection system. *Computers in Human Behavior*, *158*, 108302.

[7] Babu, A. M., & Bellamkonda, S. (2023, September). A Diagnosis Model Based on Federated Learning for Lung Cancer Classification. In *International Conference on Science, Engineering Management and Information Technology* (pp. 199-218). Cham: Springer Nature Switzerland.

[8] Choure, P., Prajapat, S., &Berwal, K. (2024, July). Federated learning approach using transfer learning architectures for lung cancer detection. In *The International Conference on Computing, Communication, Cybersecurity& AI* (pp. 387-403). Cham: Springer Nature Switzerland.

[9] Choure, P., Prajapat, S., & Berwal, K. (2024, July). Federated learning approach using transfer learning architectures for lung cancer detection. In *The International Conference on Computing, Communication, Cybersecurity & AI* (pp. 387-403). Cham: Springer Nature Switzerland.

[10] Albalawi, E., TR, M., Thakur, A., Kumar, V. V., Gupta, M., Khan, S. B., &Almusharraf, A. (2024). Integrated approach of federated learning with transfer learning for classification and diagnosis of brain tumor. *BMC medical imaging*, *24*(1), 110.

[11] Albalawi, E., TR, M., Thakur, A., Kumar, V. V., Gupta, M., Khan, S. B., &Almusharraf, A. (2024). Integrated approach of federated learning with transfer learning for classification and diagnosis of brain tumor. *BMC medical imaging*, *24*(1), 110.

[12] Mehta, S., & Saini, R. (2024, November). A Secure and Collaborative Approach to Breast Tumor Detection Using Federated Learning and CNNs. In *2024 International Conference on Intelligent Computing and*

Sustainable Innovations in Technology (IC-SIT) (pp. 1-5). IEEE.

[13] Subramanian, M., Rajasekar, V., VE, S., Shanmugavadivel, K., &Nandhini, P. S. (2022). Effectiveness of decentralized federated learning algorithms in healthcare: a case study on cancer classification. *Electronics*, *11*(24), 4117.

[14] Almufareh, M. F., Tariq, N., Humayun, M., & Almas, B. (2023, December). A federated learning approach to breast cancer prediction in a collaborative learning framework. In *Healthcare* (Vol. 11, No. 24, p. 3185). MDPI.

[15] Upreti, D., Yang, E., Kim, H., &Seo, C. (2024). A Comprehensive Survey on Federated Learning in the Healthcare Area: Concept and Applications. *CMES-Computer Modeling in Engineering & Sciences*, *140*(3).