

RESEARCH

Open Access



Predicting treatment response in multicenter non-small cell lung cancer patients based on federated learning

Yuan Liu¹, Jinzao Huang^{3,4}, Jyh-Cheng Chen^{4,5}, Wei Chen¹, Yuteng Pan¹ and Jianfeng Qiu^{2*}

Abstract

Background Multicenter non-small cell lung cancer (NSCLC) patient data is information-rich. However, its direct integration becomes exceptionally challenging due to constraints involving different healthcare organizations and regulations. Traditional centralized machine learning methods require centralizing these sensitive medical data for training, posing risks of patient privacy leakage and data security issues. In this context, federated learning (FL) has attracted much attention as a distributed machine learning framework. It effectively addresses this contradiction by preserving data locally, conducting local model training, and aggregating model parameters. This approach enables the utilization of multicenter data with maximum benefit while ensuring privacy safeguards. Based on pre-radiotherapy planning target volume images of NSCLC patients, a multicenter treatment response prediction model is designed by FL for predicting the probability of remission of NSCLC patients. This approach ensures medical data privacy, high prediction accuracy and computing efficiency, offering valuable insights for clinical decision-making.

Methods We retrospectively collected CT images from 245 NSCLC patients undergoing chemotherapy and radiotherapy (CRT) in four Chinese hospitals. In a simulation environment, we compared the performance of the centralized deep learning (DL) model with that of the FL model using data from two sites. Additionally, due to the unavailability of data from one hospital, we established a real-world FL model using data from three sites. Assessments were conducted using measures such as accuracy, receiver operating characteristic curve, and confusion matrices.

Results The model's prediction performance obtained using FL methods outperforms that of traditional centralized learning methods. In the comparative experiment, the DL model achieves an AUC of 0.718/0.695, while the FL model demonstrates an AUC of 0.725/0.689, with real-world FL model achieving an AUC of 0.698/0.672.

Conclusions We demonstrate that the performance of a FL predictive model, developed by combining convolutional neural networks (CNNs) with data from multiple medical centers, is comparable to that of a traditional DL model obtained through centralized training. It can efficiently predict CRT treatment response in NSCLC patients while preserving privacy.

Keywords Non-small cell lung cancer, Chemotherapy and radiotherapy, Federated learning, Treatment response

*Correspondence:

Jianfeng Qiu
jfqiu100@gmail.com

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Introduction

Non-Small Cell Lung Cancer (NSCLC), as one of the most common cancers, exhibits high incidence and mortality rates [1]. For patients ineligible for radical surgery, the combination of radiotherapy and chemotherapy represents their primary treatment option [2]. The assessment of treatment response, which relates to the quality of survival and the effectiveness of treatment, is the key to enhancing the prognosis of patients [3]. Within the same disease stage, patients exhibit varying responses to radiotherapy. Some experience tumor shrinkage, while others manifest signs of tumor progression [4]. Current criteria for Response Evaluation Criteria in Solid Tumors (RECIST) [5] judge treatment effectiveness based on tumor size, shape, and growth rate. However, these criteria usually take time to produce observable changes, making it challenging to provide real-time therapeutic feedback during the early stages [6]. Therefore, an urgent demand exists for a method capable of delivering real-time efficacy predictions to facilitate more informed treatment decisions.

With the advancement of artificial intelligence, the prediction of treatment response has become efficient and accurate within the precision medicine. Liu et al. [7] used MRI images to predict treatment response to chemotherapy in patients with nasopharyngeal carcinoma. Sammut et al. [8] constructed a model based on clinical, digital pathology, genomic and transcriptomic profiles to predict pathological complete responses in breast cancer cases. Xu et al. [9] constructed a deep learning (DL) model by CT images of NSCLC patients to predicted survival and cancer specific outcomes. However, highly accurate models require a wide variety of datasets. In practice, the medical image data held by each institution tends to be limited and fragmented, and the increasing awareness of data privacy makes multi-center data collection difficult. However, there are three main problems. To begin, training high-precision models often requires pooling data together in most studies [10]. Yet, medical image data held by each institution tends to be limited and fragmented, making data collection challenging. Additionally, the increasing awareness of data privacy makes multi-center data collection difficult [11]. Moreover, data centralization may bring the risk of data leakage.

Meanwhile, Federated Learning (FL) [12, 13] has garnered substantial interest within the medical domain as an emerging machine learning technique [14]. In multicenter medical research, data is typically siloed across various medical institutions and research centers, and sharing data across institutions becomes infeasible due to privacy and regulatory constraints. In contrast, FL presents a novel avenue. It allows multiple data holders to collaboratively construct machine learning models

without sharing raw data [15, 16]. This approach protects patient privacy by performing model training on local devices and sharing only model parameters. Pati et al. [17] employed FL to detect sub-compartment boundaries of glioblastoma. Similarly, Islam et al. [18] leveraged FL to build CNN architectures to identify brain tumors in MRI images. Likewise, Yan et al. [19] used FL for the automatic detection of COVID-19 lesions in images. However, existing research in this area has significant limitations, particularly regarding diversity. The evaluation of efficacy using FL has not been fully explored. Furthermore, there are limitations in applying FL in real-world environments. Most studies have primarily focused on experiments in simulation environment, lacking validation in actual medical settings. This limits the practical application value of FL.

The objective of this study aims to address this gap by exploring the application of FL in real-world settings for predicting the response to radiotherapy treatment in a multicenter cohort of NSCLC patients. We've set up a privacy-preserving data analysis framework through cross-institutional federated learning in collaboration with multiple medical centers. It allows medical image data from different centers to collaborate in training efficacy prediction models while ensuring data privacy. Through this study, we hope to offer novel insights into the treatment of NSCLC patients, paving the way for innovative prospects in multicenter collaborative research. Ultimately, our goal is to assist physicians in enhancing the precision of treatment response assessment and optimizing cancer treatment outcomes.

Methods

Patients

Image data of NSCLC patients with planning target volume images made within 1–3 days before radiotherapy at four hospitals from 2016 to 2022 was reviewed. The research received approval from the Ethics Committee of the Affiliated Hospital of Shandong First Medical University (No. SB-KJCX2101) and the Ethics Committee of the Xiangya Hospital of Central South University (No. 202,207,167). The requirement of written informed consent was waived.

The inclusion criteria were as follows: (I) 18 years of age or older; (II) primary NSCLC; (III) CRT treatment; (IV) A CT scan both prior to and within 5 months after completing the identical course of CRT; (IV) no history of surgical removal of the tumor for treatment.

The exclusion criteria were as follows: (I) patients with incomplete clinical or imaging information; (II) failure to complete the intended treatment planning protocol; (III) patients with no information on efficacy assessment after treatment; and (IV) patients with other primary tumors. After exclusion, the study ultimately included

245 patients who met the specified conditions. Among them, 102 cases originated from an affiliated hospital of Shandong First Medical University (Hospital A), 42 cases originated from Xiangya Hospital of Central South University (Hospital B), 32 cases were from a different affiliated hospital of Shandong First Medical University (Hospital C), and 69 cases were from Cathay General Hospital (Hospital D).

According to RECIST, the outcome of patients' treatment response was assessed by two proficient radiologists at each center, who analyzed the CT images taken prior to and after CRT administration. Patients with complete response (CR) and partial response (PR) who demonstrated a positive biological effect on treatment were categorized as responsive. Conversely, patients with stable disease (SD) and progressive disease (PD) who demonstrated a limited or negative biological effect were categorized as nonresponsive [7, 20]. The protocols of the scanning in the four hospitals are shown in Supplementary 1.

Data preprocessing

Observing the collected images of planning target volume, there were differences in the naming of segmentation target volume and endangered organs in different hospitals, we first imported the raw data file and RT struct file into 3D-slicer (4.11) [21] and manually selected the region of interest (ROI) named GTV in the list of names of RT struct file. The raw data and ROI were then converted to nii format images. Due to the differences thickness, the original Dicom and ROI were resampled to $1 \times 1 \times 1 \text{ mm}^3$ using B-spline interpolation. The ROI was converted to a 3D binary matrix and the tumor region was calculated. Statistical analysis of tumor sizes at three XYZ levels was conducted, and by referencing previous studies [22–24], a size of $64 \times 64 \times 64 \text{ mm}^3$ was

determined (Supplementary 2). Subsequently, 3D patches were cut with the center of the tumor as the origin for input to the model.

To prevent overfitting, increase the data diversity, and bolster the model's generalization capability, we perform image augmentation on the training data. The methods applied were as follows: [1] Random horizontal flip [2] Random rotation within the -15° to 15° range [3] Random cropping [4] Brightness, contrast, saturation, and hue set to 0.2.

Neural network structure

A custom 3D convolutional neural network model, conv3DNet, was used in this study. This model was designed from scratch, using a train-from-initialization approach, rather than relying on a pre-trained model structure. The model consists of three convolutional layers, three 3D Max-Pooling pooling layers, two fully-connected layers, and a Softmax layer for final regression to probability. The architecture of the model is shown in Fig. 1.

Federated learning framework

The FL part of this study was implemented using a FL framework, Flower 1.3.0 [25], which consists of three main modules: server-side, client-side, and strategy [26]. The server-side handles global aggregation, while the client-side manages local training. Within the built-in strategy module, various FL training schemes are embedded, facilitating the selection of the appropriate approach to achieve model parameter aggregation on the server. The framework contains popular FL algorithms such as FedAvg [12] and FedProx [27].

The process of federated learning is presented in Fig. 2. First, all clients (participants) wait for the server (central node) to transmit the initial parameters. After receiving the initial parameters, the clients train the model locally

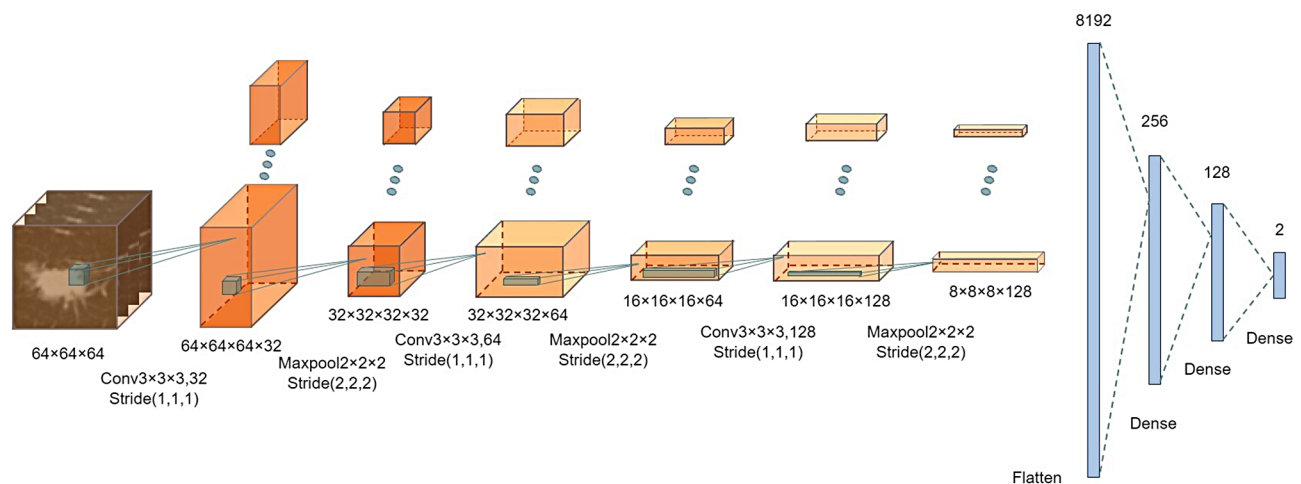


Fig. 1 The architecture of conv3DNet. Including convolution layers, max pooling layers, and fully connected layers

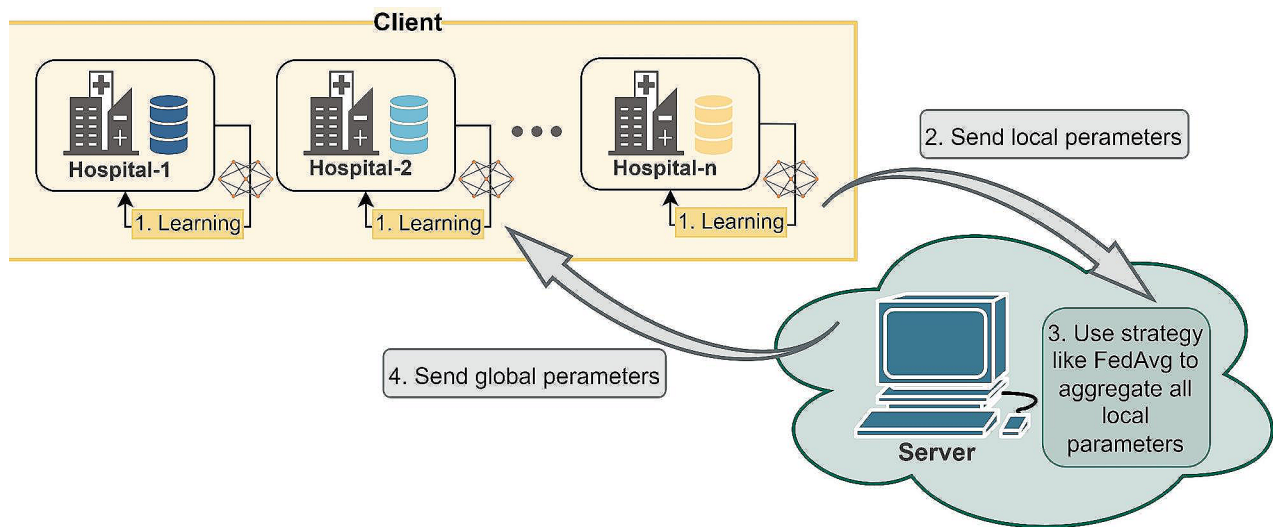


Fig. 2 The process of federated learning

using their own training set while the server remains in a waiting state. After each iteration, the client generates a model parameter update. Once a client finishes training, it transmits the model parameters to the server over the network. After receiving the parameters from all clients, the server uses the FedAvg algorithm to aggregate the parameter updates by taking their weighted average. The server then sends the aggregated parameters back to the clients over the network, and the clients train their local model again based on these aggregated parameters. The whole process is iterated repeatedly, with each iteration training the local model on the client, generating parameter updates, transmitting them to the server for aggregation, and obtaining the final global model [15, 28].

Model construction

In this study, we first built a DL model and a two-client FL model in a simulated environment, comparing their performance to elucidate the advantages of FL. Due to the unavailability of data from hospital D, we conducted a real-world three-client FL model to explore the potential of FL in healthcare applications. The model frameworks all come from conv3Dnet mentioned in 2.3.

Federated learning in a simulation environment

Initially, we developed a centralized DL model by dividing the data from hospitals A and B into training and validation sets in a ratio of 7:3. In addition, data from hospital C was used as an external validation set for testing the model, and was not involved in model training or debugging. The training hyperparameters included the following: (I) Batch size set at 8; (II) Learning rate of 0.001; (III) Utilizing the Adam optimizer; and (IV) Training for 100 epochs. The loss function employed was cross entropy.

The data were then utilized to build a FL model. Two clients (Hospital A and B) divided the local data into a training set and a validation set in a 7:3 ratio. The data from Hospital C was used as an external validation set for model testing and was not involved in model training or debugging. We configured the initial global model for both clients and used the Flower framework to communicate with the central server via gRPC to build the FL model. The training hyperparameters included the following: (I) Batch size set at 16; (II) Learning rate of 0.001; (III) Utilizing the SGD optimizer (To assess the initial performance of the FL model, we employ an SGD optimizer capable of fine-tuning the model.); (IV) Conducting 10 communication rounds; and (V) Training for 50 local epoch per client. (Local epoch means that each client trains with its local data before sending model parameters to center server.) The loss function used remained cross entropy.

Federated Learning in Real-world environments

The experiment in FL in the real world was continued using the Flower framework. The configuration of the Flower framework to process data from different healthcare organizations and protect data privacy is described. The approach is as follows: Three clients (hospital A, hospital B and hospital D) divided the local data into a training set and a validation set in a 7:3 ratio. The data from hospital C was used as an external validation set for model testing and was not involved in model training and debugging. The model training is all performed on the client side, and the flow is shown in Fig. 3. The training hyper-parameters were as follows: (I) Batch size set at 16 (II) Learning rate of 0.001; (III) Utilizing the Adam optimizer (In real-world environment with higher data distribution and complexity, we use the Adam optimizer,

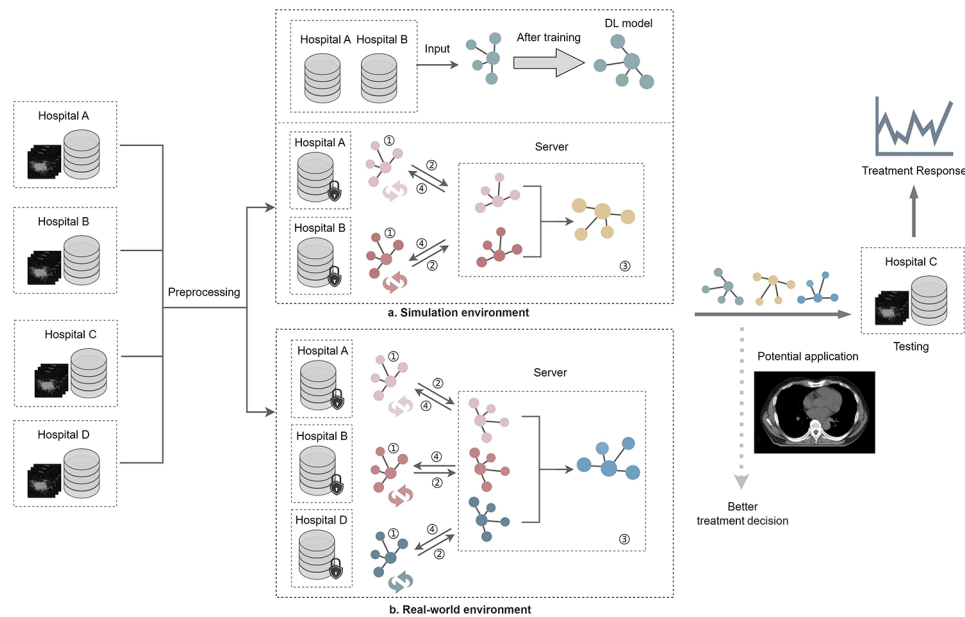


Fig. 3 Model training architecture for NSCLC treatment response from CT images. (FL training process: ①Model is trained using local data. ②Model parameters are sent to the server. ③Server aggregation parameters. ④Update parameters.)

which adaptively adjusts the learning rate for faster and more stable convergence.); (IV) Conducting 10 communication rounds; and (V) Training for 50 local epoch per client. The loss function employed remained cross entropy.

In deep learning, the model's performance is influenced by the distribution of the dataset. To assess the robustness of the proposed model, a switch to a different dataset for external validation was made, following the method described earlier. In the simulation environment, we used data from hospitals A and C to train the centralized learning model DL2, while two clients (Hospital A and C) were used to train the distributed learning model FL3. In the real world, three clients (Hospital A, C, and D) were used to train the distributed learning model FL4. Experimental parameters are shown in Supplementary 3. Data from Hospital B was used as an external validation set to test models DL2, FL3, and FL4, and was not involved in model training or debugging.

Experimental environment

Statistical analyses were performed using IBM SPSS Statistics 21.0 software. CPU processor is Intel(R) Core(TM) i5-11500 @2.70 GHz 2.71 GHz, Intel(R) Xeon(R) Silver 4114 CPU @ 2.20 GHz 2.19 GHz. The GPU processor is NVIDIA Quadro P4000; the operating system is 64 bit Windows 10 Professional with python 3.7; and the network model is implemented using a deep learning framework based on Pytorch(1.13.1).

Results

To train a multicenter well-performing predictive model, a highly diverse dataset is essential. In light of this, we collected data from a total of 245 NSCLC patients from four centers. Among these, 110 cases were classified within the responsive group, while the remaining 135 cases fell into the non-responsive group. A detailed summary of the cohort's demographic information is presented in Table 1. To ensure no significant differences in patient characteristics across institutions, we conducted statistical analyses using one-way ANOVA, Chi-square tests, and Fisher's exact test. Results indicated that, aside from gender, all other characteristics were not significantly different across institutions ($p > 0.05$). Although gender differences among the four cohorts were significant ($p < 0.05$), we believe gender does not impact our results as our analysis primarily relies on image features.

The performance of centralized DL models (DL1, DL2) and FL models (FL1, FL3) in simulated environments, as well as FL models (FL2, FL4) in the real world, are evaluated using accuracy, specificity, AUC value, and confusion matrix. For a fair comparison, the models are all evaluated using the same test set (Hospital B or C).

When validated using the dataset from hospital C in the simulated environment, the centralized DL1 model exhibits an AUC value of 0.718(95% CI: 0.52–0.88). The FL1 model achieves higher AUC value of 0.725(95% CI: 0.55–0.90). In the real-world setting, the FL2 model shows an AUC value of 0.698(95% CI: 0.49–0.87). After switching dataset B for external validation, the DL2 model built in the simulated environment achieved an AUC value of 0.695(95% CI: 0.45–0.90), while the FL3

Table 1 Patient characteristics

Subjects	Hospital A (n = 102)	Hospital B (n = 42)	Hospital C (n = 32)	Hospital D (n = 69)	P value
R	51	31	17	11	
nR	51	11	15	58	
Gender					<i>P</i> < 0.05
Male	82	34	23	36	
Female	20	8	9	33	
Age	64.32 ± 7.81	64.21 ± 10.68	63.96 ± 8.57	68.7 ± 11.4	0.057
Range	47–81	44–83	39–79	37–91	
Histological type					0.269
LUAD	48	17	14	40	
LUSC	54	25	18	29	
Tumor stage					0.120
T1	10	11	3	8	
T2	32	16	10	21	
T3	34	7	11	16	
T4	26	8	8	24	
Node stage					0.114
N0	16	10	7	20	
N1	9	5	6	7	
N2	51	13	13	19	
N3	26	14	6	23	
Metastasis					0.065
M0	66	24	22	32	
M1	36	18	10	37	
Clinical stage					0.187
IIIA	26	7	12	12	
IIIB	36	14	14	26	
IIIC	7	3	2	4	
IV	33	18	4	27	

R, responsive; nR, non-responsive. LUAD, Lung adenocarcinoma; LUSC, Lung squamous cell carcinoma

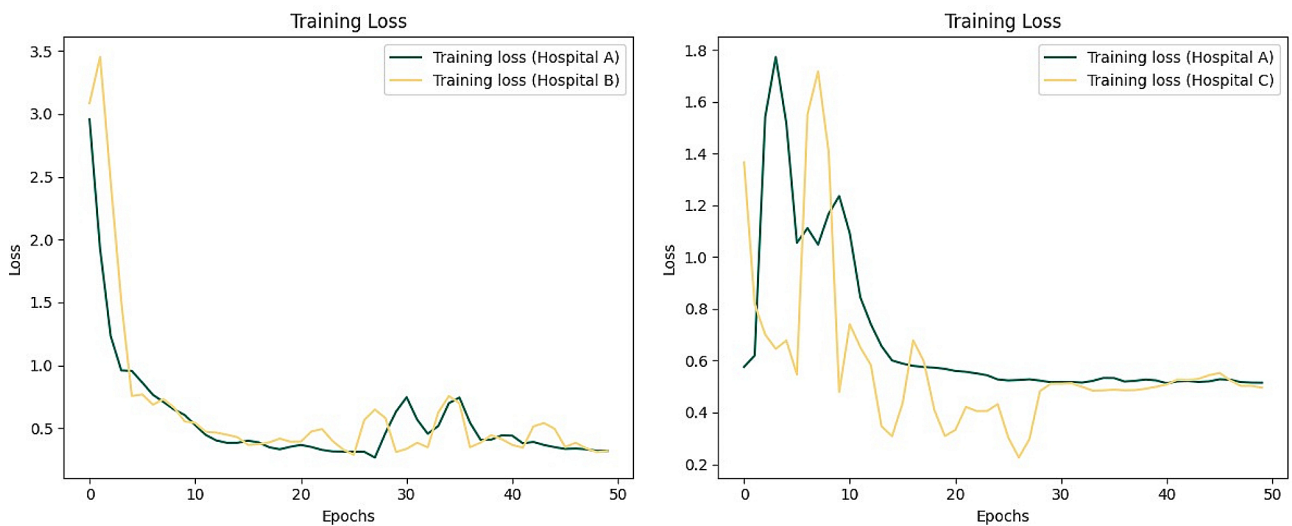


Fig. 4 The training loss of the FL model across the two hospitals

model achieved an AUC value of 0.689(95% CI: 0.51–0.85). Additionally, the FL4 model built in the real world attained an AUC of 0.672(95% CI: 0.45–0.89). Figure 4 summarizes the training loss of the FL model across

the two hospitals. Figures 5 and 6 summarizes the ROC curves and confusion matrices obtained from the above three models. Tables 2 and 3 summarize the performance metrics for our model’s training, validation, and testing.

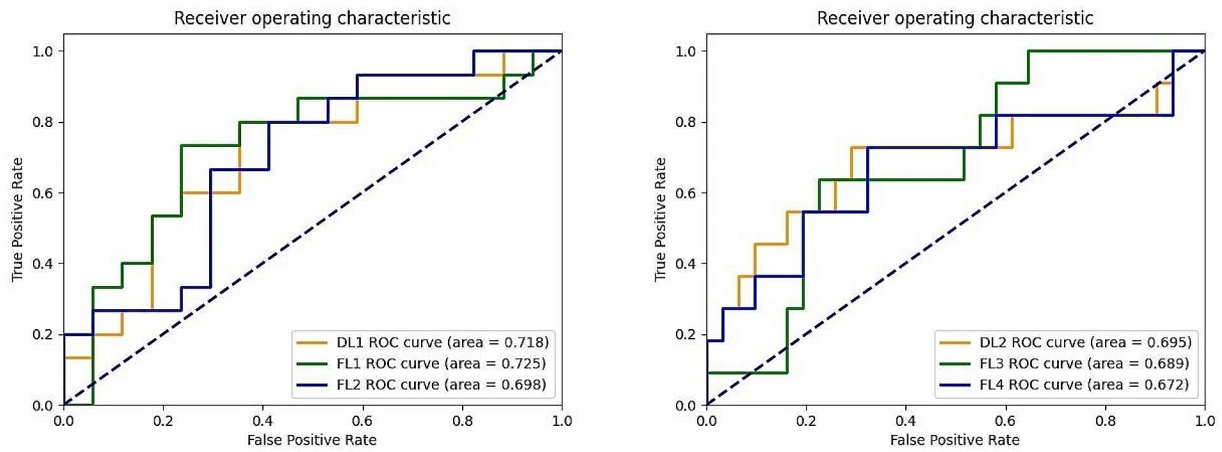


Fig. 5 Performance comparison between local and collaborative FL training based on imaging data to predict treatment response in NSCLC patients. ROC curve of three models

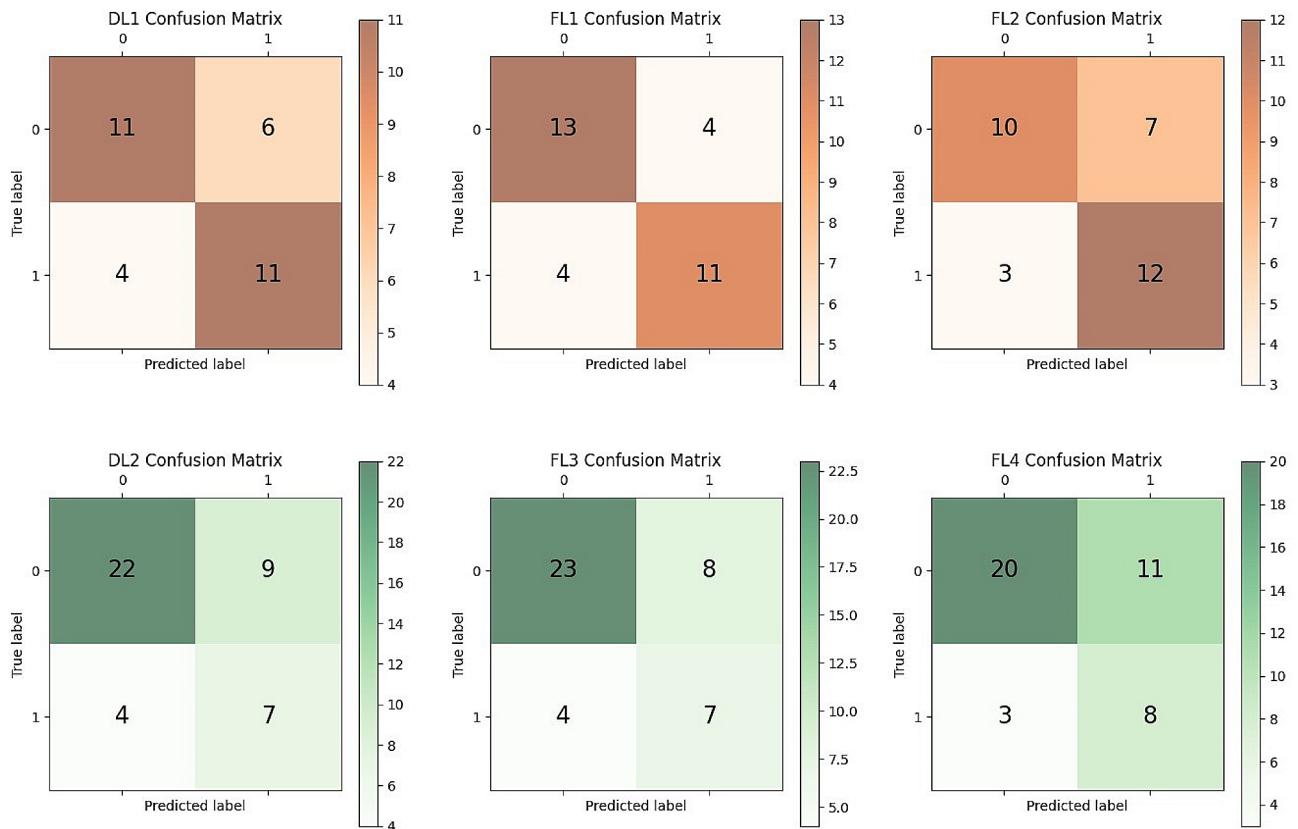


Fig. 6 Performance comparison between local and collaborative FL training based on imaging data to predict treatment response in NSCLC patients. Confusion matrix of three models

Discussion

In previous multicenter studies, machine learning or deep learning methods are usually employed to construct models using diverse medical imaging data such as CT and MRI. For instance, Cui et al. [29] developed a DL model for predicting individual patient responses

to neoadjuvant chemotherapy based on CT images of patients with locally progressed gastric cancer from four hospitals in China. Braman et al. [30] built a machine learning model to predict the ability of neoadjuvant chemotherapy to provide a complete remission of the pathology through MRI images of breast cancer patients.

Table 2 Three models performance on treatment response prediction using hospital_C test set

	Training			Validation			Testing					
	Acc	AUC	Spe	Recall	Acc	AUC	Spe	Recall	Acc	AUC	Spe	Recall
DL1	0.760	0.807	0.789	0.721	0.682	0.720	0.720	0.632	0.688	0.718	0.647	0.733
FL1	0.796	0.761	0.823	0.755	0.718	0.740	0.709	0.674	0.750	0.725	0.765	0.733
FL2	0.734	0.792	0.663	0.762	0.667	0.693	0.669	0.683	0.688	0.698	0.588	0.800

Acc: Accuracy Spe: Specificity. DL1: Centralized deep learning model trained with AB hospital data in simulation environment. FL1: Distributed federated learning model trained with AB hospital data in simulation environment. FL2: Distributed federated learning model trained with ABD hospital data in real-world environment.

However, these conventional methods typically necessitate data centralization for training, raising concerns about data privacy.

Therefore, methods such as swarm learning (SL) [31] and FL allow for training models in multi-center collaborations without sharing sensitive data. SL is a decentralized machine learning approach that does not require server-coordinated parameters, enabling direct communication between parties through a blockchain network. For instance, Saldanha et al. [32] used SL in a multicenter study to predict gene mutation status and microsatellite instability. Another study employed SL to predict molecular biomarkers for gastric cancer [33], both of which achieved remarkable results.

Unlike SL, FL utilizes a central server for coordination, through which all participants communicate. This method effectively integrates multicenter data while protecting data privacy. For example, Sheller et al. [34] divided the BraTs dataset into 10 simulated institutions to study simulated FL, which aimed to distinguish healthy brain tissue from cancerous tissue. Sadilek et al. [35] conducted several studies on FL in different scenarios to explore its performance. However, many of the FL studies in existing research have been conducted in simulated environments. These studies primarily focus on technological innovations in data security and privacy protection, but lack validation in real healthcare environments. In contrast, our study focuses more on predicting efficacy in practical clinical applications and validates the feasibility of FL in real clinical settings.

In this study, we introduced the Flower FL framework to establish a collaborative multicenter learning model based on 3D CT images for predicting the treatment response of radiotherapy in NSCLC patients. Our research involved a cohort of 245 patients from four different hospitals. We began with a theoretical performance comparison conducted in a simulation environment using data from three of these hospitals. By comparing the performance of a DL model built by a centralized approach with a FL model, our findings support the effectiveness of the FL approach [DL model accuracy=0.688/0.691, AUC=0.718/0.695; FL model accuracy=0.750/0.714, AUC=0.725/0.689]. In real-world scenarios, in order to more closely match the actual medical application scenarios and to address the challenges of data acquisition and privacy protection, we utilize data from all four hospitals to develop a FL model [FL model accuracy=0.688/0.667; area under the curve (AUC)=0.698/0.672].

The left panel of Fig. 4 shows that after 20 epochs, the loss values fluctuate. And the right panel shows that as training proceeds, the losses for both hospitals gradually decrease and stabilize. The relatively smoother loss curve for Hospital A may indicate a larger amount of data at

Table 3 Three models performance on treatment response prediction using hospital B test set

	Training			Validation			Testing					
	Acc	AUC	Spe	Recall	Acc	AUC	Spe	Recall	Acc	AUC	Spe	Recall
DL2	0.741	0.771	0.723	0.761	0.732	0.724	0.714	0.750	0.691	0.695	0.710	0.636
FL3	0.736	0.777	0.740	0.733	0.693	0.680	0.683	0.712	0.714	0.689	0.742	0.636
FL4	0.720	0.749	0.657	0.763	0.666	0.705	0.641	0.736	0.667	0.672	0.645	0.727

DL2: Centralized deep learning model trained with AC hospital data in simulation environment. FL3: Distributed federated learning model trained with AC hospital data in simulation environment. FL4: Distributed federated learning model trained with AC hospital data in real-world environment

this site, facilitating smoother learning for the model [36]. As demonstrated in Tables 2 and 3; Figs. 5 and 6, since the model weights received by the federated global model are the weighted average of the local model weights from other clients, which are aggregated by the global model and then returned to the client's local model, each client can benefit from the experiences of the other clients, thus the FL1, FL3 model's performance, surpasses that of the DL1, DL2 model. The federated learning models FL2 and FL4, trained using data from three clients, exhibited lower metrics compared to FL1 and FL3, which were trained using data from two clients. This discrepancy may attribute to uneven data distribution in hospital D compared to the other two clients. In cases where data distribution differs noticeably among clients, there may be inconsistencies in data distribution during model aggregation. Using fewer clients reduces the likelihood of this data distribution inconsistency. In addition, the performance of the model trained using AC hospital data (with hospital B as the test set) is comparatively lower than that trained using AB hospital data (with hospital C as the test set). In FL, hospital A, which possesses a larger dataset, is assigned a higher weight. Hence, this discrepancy in performance could potentially be attributed to the fact that the data distribution of Hospital C aligns more closely with that of Hospital A compared to Hospital B.

Certain limitations remain in this study. First, the dataset of NSCLC patients used in this paper was relatively small, and the inclusion of data from more medical institutions could potentially enhance the model's performance. In addition, this study was limited by the sample size, and CR and PR were categorized as treatment with remission, and SD and PD were categorized as treatment without remission for dichotomous studies. In the future, it is hoped that the RECIST criteria will be used to classify efficacy into four classes for more accurate prediction, while expanding and balancing the sample size. Furthermore, this study focused on constructing a CT-based unimodal model using FL, omitting the integration of additional data such as clinical features and pathology image features, which have the potential to enhance the model's predictive capacity in the context of cancer treatment response. Future studies will aim to expand the size of the dataset, invite more medical institutions to participate, and integrate data from various sources to build a more comprehensive model for the precise prediction of radiotherapy treatment response in NSCLC patients.

Conclusions

To emphasize the efficacy of a distributed learning approach in a data-private setting, we conducted a study on FL for predicting treatment responses among patients with NSCLC.

We compared traditional DL and FL approaches. Our results show that FL can achieve comparable performance to centralized DL without sharing sensitive data. In addition, we validate the feasibility of FL in real-world applications. We believe that this approach is not only applicable to NSCLC efficacy prediction, but can also be extended to other DL applications for medical image analysis. This research provides an effective approach to address data privacy and collaboration issues in multi-center medical image analysis, which is expected to have a broader impact in clinical applications.

Abbreviations

NSCLC	Non-small cell lung cancer
FL	Federated learning
SL	Swarm learning
CRT	Chemotherapy and radiotherapy
DL	Deep learning
CNNs	Convolutional neural networks
RECIST	Response Evaluation Criteria in Solid Tumors
CR	Complete response
PR	Partial response
SD	Stable disease
PD	Progressive disease

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12885-024-12456-7>.

Supplementary Material 1

Supplementary Material 2

Supplementary Material 3

Acknowledgements

Not applicable.

Author contributions

JQ conceived of the presented idea and supervised the project. YL, JH and WC conducted image preprocessing and analysis. YL and YP collected the data. YL wrote the manuscript. JQ and JC reviewed the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by China National Key Research and Development (No. 2021YFE0204600), the Academic Promotion Program of Shandong First Medical University (No. 2019QL009).

Data availability

The data that support the findings of this study are not openly available due to reasons of sensitivity and are available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate

The research received approval from the Ethics Committee of the Affiliated Hospital of Shandong First Medical University (No. SB-KJXC2101) and the Ethics Committee of the Xiangya Hospital of Central South University (No. 202207167). The requirement of written informed consent was waived.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹School of Radiology, Shandong First Medical University and Shandong Academy of Medical Sciences, Taian, China

²School of Radiology, Second Affiliated Hospital of Shandong First Medical University and Shandong Academy of Medical Sciences, Taian, China

³Department of Radiology, Cathay General Hospital, Taipei, China

⁴Department of Biomedical Imaging and Radiological Sciences, National Yang-Ming Chiao-Tung University, Taipei, China

⁵Department of Biomedical Imaging and Radiological Science, China Medical University, Taichung, China

Received: 24 April 2024 / Accepted: 30 May 2024

Published online: 05 June 2024

References

- Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal AJCacjfc. Global cancer statistics, 2012. *CA Cancer J Clin*. 2015;65(2):87–108.
- Miller M, Hanna NJB. Advances in systemic therapy for non-small cell lung cancer. *BMJ*. 2021;375.
- Detterbeck FC, Boffa DJ, Kim AW, Tanoue LTJC. The eighth edition lung cancer stage classification. *Chest*. 2017;151(1):193–203.
- Wang L, Zhan C, Gu J, Xi J, Lin Z, Xue L, et al. Role of skip mediastinal lymph node metastasis for patients with resectable non-small-cell lung cancer: a propensity score matching analysis. *Clin Lung Cancer*. 2019;20(3):e346–55.
- Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer*. 2009;45(2):228–47.
- Suzuki C, Jacobsson H, Hatschek T, Torkzad MR, Bodén K, Eriksson-Alm Y, et al. Radiologic measurements of tumor response to treatment: practical approaches and limitations. *Radiographics*. 2008;28(2):329–44.
- Liu J, Mao Y, Li Z, Zhang D, Zhang Z, Hao S, et al. Use of texture analysis based on contrast-enhanced MRI to predict treatment response to chemoradiotherapy in nasopharyngeal carcinoma. *J Magn Reson Imaging*. 2016;44(2):445–55.
- Sammur S-J, Crispin-Ortuzar M, Chin S-F, Provenzano E, Bardwell HA, Ma W, et al. Multi-omic machine learning predictor of breast cancer therapy response. *Nature*. 2022;601(7894):623–9.
- Xu Y, Hosny A, Zeleznik R, Parmar C, Coroller T, Franco I, et al. Deep learning predicts lung cancer treatment response from serial medical imaging. *Clin Cancer Res*. 2019;25(11):3266–75.
- Willemink MJ, Koszek WA, Hardell C, Wu J, Fleischmann D, Harvey H, et al. Preparing medical imaging data for machine learning. *Radiology*. 2020;295(1):4–15.
- Kairouz P, McMahan HB, Avent B, Bellet A, Bennis M, Bhagoji AN, et al. Advances and open problems in federated learning. *Found trends® Mach Learn*. 2021;14(1–2):1–210.
- McMahan B, Moore E, Ramage D, Hampson S, y, Arcas BA, editors. Communication-efficient learning of deep networks from decentralized data. *Artificial intelligence and statistics*; 2017: PMLR.
- Li T, Sahu AK, Talwalkar A, Smith VJL. Federated learning: challenges, methods, and future directions. *IEEE Signal Process Mag*. 2020;37(3):50–60.
- Sarma KV, Harmon S, Sanford T, Roth HR, Xu Z, Tetreault J, et al. Federated learning improves site performance in multicenter deep learning without data sharing. *J Am Med Inform Assoc*. 2021;28(6):1259–64.
- Yang Q, Liu Y, Chen T, Tong, YJAToLS. Technology. Federated machine learning: Concept and applications. *ACM Trans Intell Syst Technol (TIST)*. 2019;10(2):1–19.
- Rieke N, Hancox J, Li W, Milletari F, Roth HR, Albarqouni S, et al. The future of digital health with federated learning. *NPJ Digit Med*. 2020;3(1):1–7.
- Pati S, Baid U, Edwards B, Sheller M, Wang S-H, Reina GA, et al. Federated learning enables big data for rare cancer boundary detection. *Nat Commun*. 2022;13(1):7346.
- Islam M, Reza MT, Kaosar M, Parvez MZJNPL. Effectiveness of federated learning and CNN ensemble architectures for identifying brain tumors using MRI images. *Neural Process Lett*. 2023;55(4):3779–809.

19. Yan B, Wang J, Cheng J, Zhou Y, Zhang Y, Yang Y, et al. editors. Experiments of federated learning for COVID-19 chest X-ray images. *Advances in Artificial Intelligence and Security: 7th International Conference, ICAIS 2021, Dublin, Ireland, July 19–23, 2021, Proceedings, Part II 7*; 2021: Springer.
20. Yang F, Zhang J, Zhou L, Xia W, Zhang R, Wei H et al. CT-based radiomics signatures can predict the tumor response of non-small cell lung cancer patients treated with first-line chemotherapy and targeted therapy. *Eur Radiol.* 2022;1–10.
21. Fedorov A, Beichel R, Kalpathy-Cramer J, Finet J, Fillion-Robin J-C, Pujol S, et al. 3D slicer as an image computing platform for the quantitative Imaging Network. *Magn Reson Imaging.* 2012;30(9):1323–41.
22. Liang Y, Lee K, Bovi JA, Palmer JD, Brown PD, Gondi V, et al. Deep learning-based automatic detection of brain metastases in heterogenous multi-institutional magnetic resonance imaging sets: an exploratory analysis of NRG-CC001. *Int J Radiation Oncology* Biology* Phys.* 2022;114(3):529–36.
23. Huang W, Xue Y, Wu YJP. A CAD system for pulmonary nodule prediction based on deep three-dimensional convolutional neural networks and ensemble learning. *PLoS ONE.* 2019;14(7):e0219369.
24. Zhuge Y, Ning H, Mathen P, Cheng JY, Krauze AV, Camphausen K, et al. Automated glioma grading on conventional MRI images using deep convolutional neural networks. *Med Phys.* 2020;47(7):3044–53.
25. Beutel DJ, Topal T, Mathur A, Qiu X, Fernandez-Marques J, Gao Y et al. Flower: a friendly federated learning research framework. *arXiv Preprint arXiv:200714390.* 2020.
26. Petrangeli E, Tonello N, Vallati CJ. Performance evaluation of federated learning for residential energy forecasting. *IoT.* 2022;3(3):381–97.
27. Li T, Sahu AK, Zaheer M, Sanjabi M, Talwalkar A, Smith, VJPoMI, et al. Federated optimization in heterogeneous networks. *Federated Optim Heterogen Networks.* 2020;2:429–50.
28. Darzidehkalani E, Van Ghasemi-Rad M. Federated learning in medical imaging: part II: methods, challenges, and considerations. *J Am Coll Radiol.* 2022;19(8):975–82.
29. Cui Y, Zhang J, Li Z, Wei K, Lei Y, Ren J et al. A CT-based deep learning radiomics nomogram for predicting the response to neoadjuvant chemotherapy in patients with locally advanced gastric cancer: a multicenter cohort study. *EClinicalMedicine.* 2022;46.
30. Braman NM, Etesami M, Prasanna P, Dubchuk C, Gilmore H, Tiwari P, et al. Intratumoral and peritumoral radiomics for the pretreatment prediction of pathological complete response to neoadjuvant chemotherapy based on breast DCE-MRI. *Breast Cancer Res.* 2017;19:1–14.
31. Warnat-Herresthal S, Schultze H, Shastry KL, Manamohan S, Mukherjee S, Garg V, et al. Swarm learning for decentralized and confidential clinical machine learning. *Nature.* 2021;594(7862):265–70.
32. Saldanha OL, Quirke P, West NP, James JA, Loughrey MB, Grabsch HI, et al. Swarm learning for decentralized artificial intelligence in cancer histopathology. *Nat Med.* 2022;28(6):1232–9.
33. Saldanha OL, Muti HS, Grabsch HI, Langer R, Dislich B, Kohlross M, et al. Direct prediction of genetic aberrations from pathology images in gastric cancer with swarm learning. *Gastric Cancer.* 2023;26(2):264–74.
34. Sheller MJ, Edwards B, Reina GA, Martin J, Pati S, Kotrotsou A, et al. Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data. *Sci Rep.* 2020;10(1):12598.
35. Sadilek A, Liu L, Nguyen D, Kamruzzaman M, Serghiou S, Rader B, et al. Privacy-first health research with federated learning. *NPJ Digit Med.* 2021;4(1):132.
36. Heidari A, Javaheri D, Toumaj S, Navimipour NJ, Rezaei M, Unal MJAIM. A new lung cancer detection method based on the chest CT images using Federated Learning and blockchain systems. *Artif Intell Med.* 2023;141:102572.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.