Check for updates

Received: 5 October 2023 | Revised: 10 January 2024 | Accepted: 21 January 2024

DOI: 10.1002/mp.16964

RESEARCH ARTICLE

MEDICAL PHYSICS

Differential privacy preserved federated learning for prognostic modeling in COVID-19 patients using large multi-institutional chest CT dataset

Isaac Shiri[1] | **Yazdan Salimi**[1] | **Nasim Sirjani**[2] | **Behrooz Razeghi**[3] | **Sara Bagherieh**[4] | **Masoumeh Pakbin**[5] | **Zahra Mansouri**[1] | **Ghasem Hajianfar**[1] | **Atlas Haddadi Avval**[6] | **Dariush Askari**[7] | **Mohammadreza Ghasemian**[8] | **Saleh Sandoughdaran**[9] | **Ahmad Sohrabi**[10] | **Elham Sadati**[11] | **Somayeh Livani**[12] | **Pooya Iranpour**[13] | **Shahriar Kolahi**[14] | **Bardia Khosravi**[15] | **Salar Bijari**[11] | **Sahar Sayfollahi**[16] | **Mohammad Reza Atashzar**[17] | **Mohammad Hasanian**[18] | **Alireza Shahhamzeh**[19] | **Arash Teimouri**[13] | **Neda Goharpey**[20] | **Hesamaddin Shirzad-Aski**[21] | **Jalal Karimi**[22] | **Amir Reza Radmard**[23] | **Kiara Rezaei-Kalantari**[24] | **Mostafa Ghelich Oghli**[2] | **Mehrdad Oveisi**[25] | **Alireza Vafaei Sadr**[26] | **Slava Voloshynovskiy**[3] | **Habib Zaidi**[1,27,28,29]

Correspondence
Habib Zaidi, Geneva University Hospital, Division of Nuclear Medicine and Molecular Imaging, CH-1211 Geneva, Switzerland.
Email: habib.zaidi@hcuge.ch

Funding information
The Swiss National Science Foundation, Grant/Award Number: 320030_176052

Abstract

Background: Notwithstanding the encouraging results of previous studies reporting on the efficiency of deep learning (DL) in COVID-19 prognostication, clinical adoption of the developed methodology still needs to be improved. To overcome this limitation, we set out to predict the prognosis of a large multi-institutional cohort of patients with COVID-19 using a DL-based model.

Purpose: This study aimed to evaluate the performance of deep privacy-preserving federated learning (DPFL) in predicting COVID-19 outcomes using chest CT images.

Methods: After applying inclusion and exclusion criteria, 3055 patients from 19 centers, including 1599 alive and 1456 deceased, were enrolled in this study. Data from all centers were split (randomly with stratification respective to each center and class) into a training/validation set (70%/10%) and a hold-out test set (20%). For the DL model, feature extraction was performed on 2D slices, and averaging was performed at the final layer to construct a 3D model for each scan. The DensNet model was used for feature extraction. The model was developed using centralized and FL approaches. For FL, we employed DPFL approaches. Membership inference attack was also evaluated in the FL strategy. For model evaluation, different metrics were reported in the hold-out test sets. In addition, models trained in two scenarios, centralized and FL, were compared using the DeLong test for statistical differences.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.
© 2024 The Authors. *Medical Physics* published by Wiley Periodicals LLC on behalf of American Association of Physicists in Medicine.

Results: The centralized model achieved an accuracy of 0.76, while the DPFL model had an accuracy of 0.75. Both the centralized and DPFL models achieved a specificity of 0.77. The centralized model achieved a sensitivity of 0.74, while the DPFL model had a sensitivity of 0.73. A mean AUC of 0.82 and 0.81 with 95% confidence intervals of (95% CI: 0.79–0.85) and (95% CI: 0.77–0.84) were achieved by the centralized model and the DPFL model, respectively. The DeLong test did not prove statistically significant differences between the two models (p-value = 0.98). The AUC values for the inference attacks fluctuate between 0.49 and 0.51, with an average of 0.50 ± 0.003 and 95% CI for the mean AUC of 0.500 to 0.501.
Conclusion: The performance of the proposed model was comparable to centralized models while operating on large and heterogeneous multi-institutional datasets. In addition, the model was resistant to inference attacks, ensuring the privacy of shared data during the training process.

KEYWORDS
COVID-19, CT, deep learning, federated learning, privacy, prognosis

1 | INTRODUCTION

The staggering number of fatalities due to the COVID-19 pandemic and its unceasing surges highlights the necessity of developing more effective techniques for predicting the prognosis of patients.[1,2] This helps identifying patients requiring more advanced treatment and longer hospital stays.[3] Furthermore, elucidating the full picture of disease severity in COVID-19 patients could lead to non-discrimination and fair distribution of medical resources based on their condition's gravity.[4]

Numerous strategies[1,2] have been suggested to more accurately prognosticate COVID-19 patients, including chest computed tomography (CT) images.[5] Contrary to the gold-standard diagnostic method, RT-PCR, chest CT can provide physicians with more detailed and efficient information.[6,7] In addition, it can be used to help healthcare workers grasp the full extent and magnitude of disease severity.[8] However, despite all the advantages that a CT scan offers in the context of COVID-19 infection, interpreting it is still mostly subjective and may differ based on the physician or radiologist's opinion and level of expertise.[6,9] Consequently, the need to design new methods that objectively report CT findings is emphasized.[6,10] Several scoring systems have also been proposed in the literature,[6,11] including but not limited to two studies conducted by Carbonell et al.[12] and Li et al.[13] evaluated the predictive power of quantitative chest CT assessments along with some clinical biomarkers for patient prognosis. However, notwithstanding the great promise of such models, they offer limited prognostic value and lack sufficient inter-observer unanimity.[9]

Furthermore, artificial intelligence (AI) algorithms, such as machine learning (ML) and deep learning (DL), seem promising for determining COVID-19 patients' diagnosis and prognosis, as they can decode data and comprehend information in images that are not readily visible to the naked eye.[10,14,15] In this regard, previous studies have already demonstrated that ML-based algorithms can be readily applied to medical images and provide prognostic and diagnostic models.[14,16,17] In addition, several articles have researched the intersection of DL models and COVID-19 prognosis prediction, for instance, Gong et al.[18] gathered a multi-centered cohort and evaluated the predictive power of a DL-based model, which also utilized some of the patient's clinical data. Moreover, Wang et al.[19] reported a prognostic analysis using a fully automatic DL system that successfully stratified patients into two groups, namely high- and low-risk groups, with significantly different hospital stay durations.

Medical imaging, like any other medical data, contains sensitive and private information, and misuse or unauthorized access can have serious consequences. Therefore, protecting this information is paramount.[20–23] Multiple strategies can be implemented to protect patients' privacy effectively when using medical imaging.[20,21,24,25] Different guidelines and policies were initially implemented to manage access to patient's information and images, such as passwords and multi-factor authentication methods.[26] This could be followed by anonymization techniques, such as masking, pseudonymization, perturbation, or synthesis to eras and/or obfuscating the personal identifying information to ensure that even with access to the images, individual personal information remains protected.[27] Encryption is another strategy that turns data into a secure, safe, and unreadable format accessible only to individuals with the correct decryption keys, thereby safeguarding the data during transmission and storage.[28] In addition to these strategies, differential privacy (DP) and federated learning (FL), are a complement to traditional privacy protection strategies by ensuring the confidentiality of

24734209, 2024, 7, Downloaded from https://aapm.onlinelibrary.wiley.com/doi/10.1002/mp.16964 by Bibliotheque de l'Universite de Geneve, Division de l'information, Wiley Online Library on [10/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

patient data.[20,21,25] These methods preserve privacy while enabling us to use patient data to build ML and DL algorihms.[20,21,29–31]

FL allows the development of ML/DL algorithms across multiple centers without exchanging sensitive data.[20,21,32–38] However, different attacks, such as membership or adversarial attacks, could be performed during FL algorithms development,[20,21,24] which risks patients' privacy or prevents the model from converging during training.[20,21,33] FL has recently been used in the context of COVID-19 infection to achieve more generalizable findings regarding diagnosis, treatment planning, and outcome prediction.[39] Despite the promising results of previous studies, further research on larger numbers of patients is required to illuminate the true potential and feasibility of utilizing DL as a convenient tool in COVID-19 prognostication. Thus, we performed this study to predict the prognosis of a huge multi-center cohort of patients with COVID-19 using a DL-based model. We developed a decentralized DP-preserving deep federated algorithm for COVID-19 prognostication, potentially addressing the privacy issue during DL model development in a multi-center scenario and compared it with a centralized model.

2 | MATERIALS AND METHODS

2.1 | Data acquisition and inclusion/exclusion criteria

In the first place, a total of 5940 patients from 19 centers were included in this study.[14,40] Subjects were RT-PCR test confirmed as COVID-19. Figure 1 depicts the inclusion and exclusion criteria. The obtained data from the aforementioned centers included the following items: (i) Patients' chest CT images, (ii) RT-PCR results, and (iii) the results of a 4-month follow-up for outpatient centers or at-discharge follow-up for inpatient ones based on which the subjects were sub-grouped into "alive" and "deceased".[14] The exclusion criteria in the present study were as follows: (i) Patients not followed up or transferred to another hospital (n = 1316), (ii) Cases with a confirmed diagnosis of lung cancer at any stage (n = 200), (iii) Subjects whose CT images sustained motion artifacts (n = 215) or incomplete lung presentation (n = 98) or low quality (n = 178), (iv) Patients who had negative RT-PCR (n = 860), and (v) Cases with only contrast-enhanced CT images (n = 18).[14] After excluding the 2885 cases, 3055 patients remained, including 1599 alive and 1456 deceased. One of the 19 centers was outpatient, while the remaining were inpatients. Regarding the single center which offered outpatient care, patients were medically treated based on the standard COVID-19 regimen provided by Iran's national guidelines [corona.behdasht.gov.ir].[14] In the inpatient centers, patients were discharged provided they had stable blood pressure (systolic blood pressure over 90), were not tachycardic, had an O2 saturation level of 94% or above, and did not have a fever for at least two consecutive days before discharge.[14]

2.2 | CT image acquisition

CT image acquisition was consistent with the national guidelines on high-resolution techniques.[14,41] Patients were instructed to hold their breath when scanning to avert any motion artifact.[14] Some inconsistencies were noted within the centers' acquisition parameters and techniques, including radiation dose, slice thickness, and tube current.[14] CT Dose Index ($CTDI_{vol}$) was presented as the inter/intra-centric variation parameter.[14] Table 1 summarizes the mentioned CT acquisition parameters for each center, along with the number of CT images, alive and deceased cases, and percentage of male and female. In all centers, each CT image was interpreted by two experienced radiologists, and the image report was submitted unanimously based on the COVID-19 reporting and data system (CO-RADS) guideline.[6,14] In case of disagreement between the reports of the first two radiologists, a third experienced radiologist was asked to read the images and prepare the final report.[14]

2.3 | Image segmentation, preprocessing, and data splitting

An automatic DL-based "COLI-Net" model was utilized to segment the lungs.[42] The images were cropped based on the lung mask and resized to 296 × 296 for the model to perform efficacious computations. This segmentation was not used for further model training. Data from all centers was split (randomly with stratification respective to each center and class) into a training/validation set (70%/10%) and a hold-out test set (20%). All evaluations and metrics reported were performed on the hold-out test set.

2.4 | Deep learning model

As the different centers use dissimilar image acquisition protocols, and the slice thickness was highly variable across the different scans, we employed a combined 2D-3D model for the DL model core. Therefore, feature extraction was performed in 2D slices, and averaging was performed at the final layer during training to construct a 3D model for each scan. The DensNet[43] model was used for 2D feature extraction. The training was performed using the Adam optimizer. An initial learning rate of 10^{-5} and a decay of 0.05 for 300 iterations were used. A summary of the network architecture is presented in Figure 2.

24734209, 2024, 7, Downloaded from https://aapm.onlinelibrary.wiley.com/doi/10.1002/mp.16964 by Bibliotheque de l'Universite de Geneve, Division de l'information, Wiley Online Library on [10/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

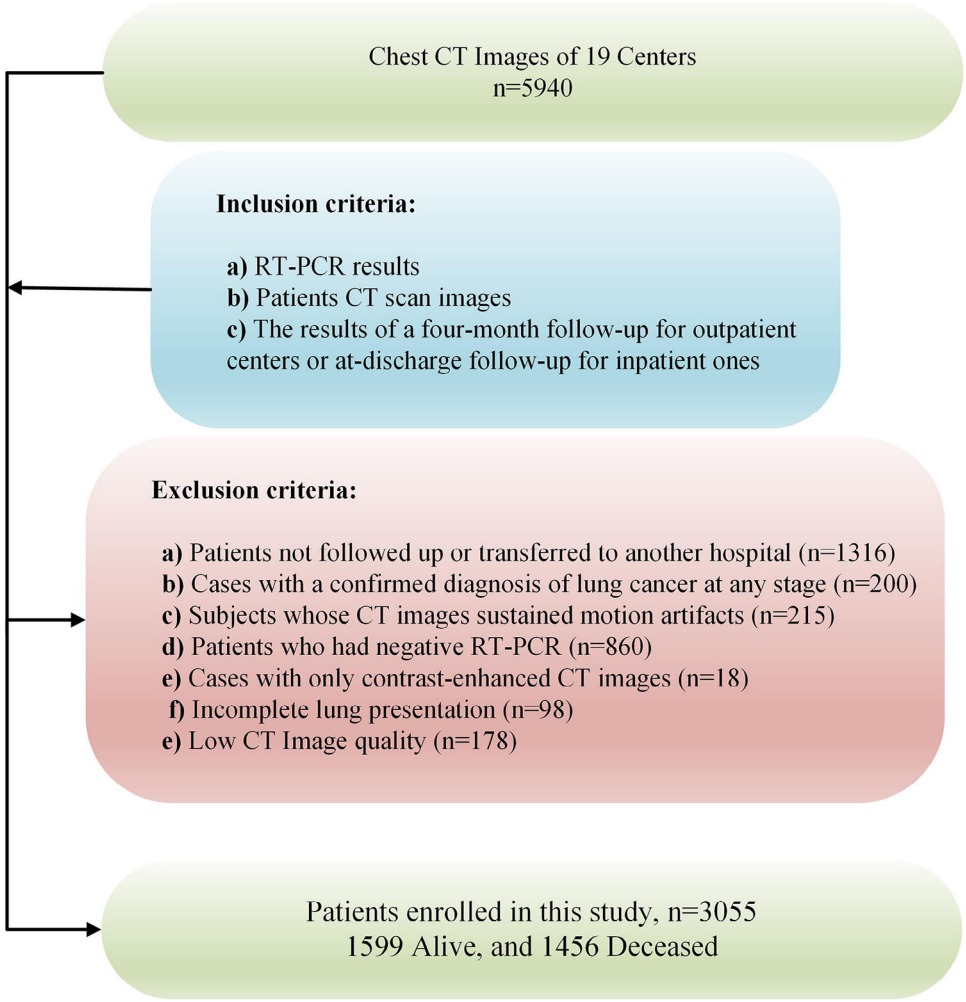

FIGURE 1 Inclusion and exclusion criteria were used in this study for data gathering. Initially, 5940 patients from 19 centers tested RT-PCR positive for COVID-19 and were considered. Inclusion criteria included chest CT images, RT-PCR results, and follow-up outcomes. Exclusions were made for patients lacking follow-up (1316), diagnosed with lung cancer (200), with CT artifacts (215), incomplete lung presentation (98), low-quality images (178), negative RT-PCR results (860), or contrast-only CTs (18). The final cohort comprised 3055 patients, of which 1599 were alive, and 1456 were deceased at follow-up.

2.5 | Training strategies

The selection of a training strategy (center-based, centralized, and FL) is affected by several variables, including the availability of a central location, data size, computational power, the necessity of moving data, and the requirement to preserve privacy.[20,21] Center-based and centralized training approaches are appropriate when a central location with a large data set and computational power is available. In contrast, FL methods are appropriate when data movement is not feasible.[44,45] We implemented centralized and FL strategies in the current study and compared their performance (Figure 2). In a centralized strategy, data are collected at multiple centers and pooled in a central location for training.[46,47] Data should move from centers in this scenario, and patients' privacy is hardly preserved, limiting the potential of acquiring data sets from multiple centers.[46,47] In FL, data are collected at various centers, and a model is trained by distributing the training process across multiple centers without the need for data movement.[46] The FL approach allows for distributed training without needing data movement or a central location.[44,45]

2.6 | Introduction to aggregation approach

The FL model aggregates updates from multiple data sources to produce a global model.[20,21,33,35] However, the arithmetic means which is widely used for aggregation, can be vulnerable to data corruption.[20,21,33,35,38] One approach to increase robustness to such corruption is to use an approximate geometric median.[20,21,33,35,38,48] Other possible

24734209, 2024, 7, Downloaded from https://aapm.onlinelibrary.wiley.com/doi/10.1002/mp.16964 by Bibliotheque de l'Universite de Geneve, Division de l'information, Wiley Online Library on [10/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 Patient demographics and CT acquisition parameters in the different centers. This table details the number of patients, survival status (alive, dead), gender distribution (female and male percentages), and CT acquisition parameters (Tube current in mAs, $CTDI_{vol}$, and slice thickness in mm) for each center.

Centers	Number	Alive	Dead	Female (%)	Male (%)	Tube current (mAs)	$CTDI_{vol}$	Slice thickness (mm)
Center 01	16	11	5	25.0	75.0	113.6 ± 100.6	7.06 ± 3.97	1.56 ± 0.99
Center 02	21	15	6	28.6	71.4	232 ± 32.8	8.21 ± 0.13	1.49 ± 0.50
Center 03	25	16	9	54.0	46.0	113.4 ± 44.9	8.15 ± 3.16	2.48 ± 2.08
Center 04	33	19	14	42.4	57.6	228.2 ± 112.7	7.61 ± 5.24	2.31 ± 1.18
Center 05	39	22	17	44.0	56.0	452.7 ± 128	10.41 ± 2.67	2.00 ± 0.00
Center 06	40	21	19	50.0	50.0	152.9 ± 41.2	6.92 ± 2.57	2.46 ± 0.24
Center 07	150	77	73	43.3	56.7	127 ± 4.8	6.31 ± 0.81	7.29 ± 1.74
Center 08	176	92	84	47.1	52.9	186.6 ± 63	13.4 ± 0.00	3.02 ± 1.23
Center 09	181	94	87	54.7	45.3	218.8 ± 38	9.06 ± 5.19	6.54 ± 1.22
Center 10	187	97	90	51.7	48.3	124.2 ± 10.4	6.07 ± 1.25	2.00 ± 0.16
Center 11	192	105	87	23.9	76.1	143.6 ± 39.7	5.19 ± 1.96	2.33 ± 0.24
Center 12	197	99	98	43.7	56.3	81.1 ± 42.3	4.56 ± 2.77	4.65 ± 1.17
Center 13	200	92	108	61.0	39.0	149.1 ± 43.9	6.24 ± 4.14	4.92 ± 0.48
Center 14	203	104	99	63.9	36.1	174.4 ± 55.9	6.69 ± 2.60	4.97 ± 0.35
Center 15	214	111	103	53.2	46.8	210.4 ± 49.1	6.27 ± 1.69	1.79 ± 1.28
Center 16	218	115	103	44.7	55.3	84.2 ± 44.7	4.82 ± 3.37	4.99 ± 0.42
Center 17	294	147	147	36.1	63.9	166.3 ± 39.3	6.10 ± 2.92	6.58 ± 1.38
Center 18	318	178	140	59.0	41.0	153.4 ± 48.5	6.51 ± 3.29	5.17 ± 0.65
Center 19	351	184	167	49.0	51.0	173.7 ± 37.9	6.21 ± 1.38	7.64 ± 1.08

approaches are zeroing and clipping techniques.[33] Zeroing involves replacing model updates that exceed a predetermined threshold with zeros, while clipping involves bounding the L2 norm of updates by projecting them onto an L2 ball of a certain radius.[33,49] The hyperparameter determines this radius and setting it too high can add too much noise, while putting it too low can cause a high bias in gradient estimation.[33]

2.7 | Mathematics of Gaussian differentially private federated averaging with adaptive quantile clipping (GDP-AQuCl) approach

DP protects the privacy of training data in ML by adding noise to model parameters, ensuring that the output distributions of a randomized algorithm are close for any neighboring inputs that differ by only one data point.[29,33,35,38] The privacy budget ϵ and the probability of information leakage δ can be adjusted to control the level of privacy protection.[33,35,38,50] In the current study, we implement the Gaussian differentially private federated weighted averaging with an adaptive quantile clipping (GDP-AQuCl) approach that combines DP with the adaptive quantile clipping method to provide strong privacy and performance guarantees for FL.[33,35,38,51] In this approach, Gaussian noise is added to model updates, and the updates are averaged using the adaptive quantile clipping method, which excludes outlier models and ensures that the global model is not overly influenced by a small number of poorly performing models.[33,35,38,52] This combination of techniques helps protecting individual data sources' privacy while still producing a high-quality global model.[33,35,38,53]

In the GDP-AQuCl approach, the global model parameters at iteration $t+1$ are updated using the adaptive quantile clipping function and the DP function.[33,35,53] The global model parameters are updated by taking the sum of the model updates from all decentralized data and applying the adaptive quantile clipping function to each update to ensure that the L_2 norm of the updates is bounded, and then adding noise uses the DP function to protect the privacy of the data used to compute the updates.[33,35,53] The updated global model parameters are then given by:

$$\theta_{t+1} = \theta_t + \frac{1}{n}\sum_{i=1}^{n} Clip\left(DP\left(\theta_t - \theta_i\right), C_t\right)$$

where θ_t and θ_{t+1} are the global model parameters at iterations t and $t+1$, respectively, n is the number of decentralized data sources, $Clip(., C_t)$ is the adaptive quantile clipping function and $DP(.)$ is the DP

24734209, 2024, 7, Downloaded from https://aapm.onlinelibrary.wiley.com/doi/10.1002/mp.16964 by Bibliotheque de l'Universite de Geneve, Division de l'information, Wiley Online Library on [10/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

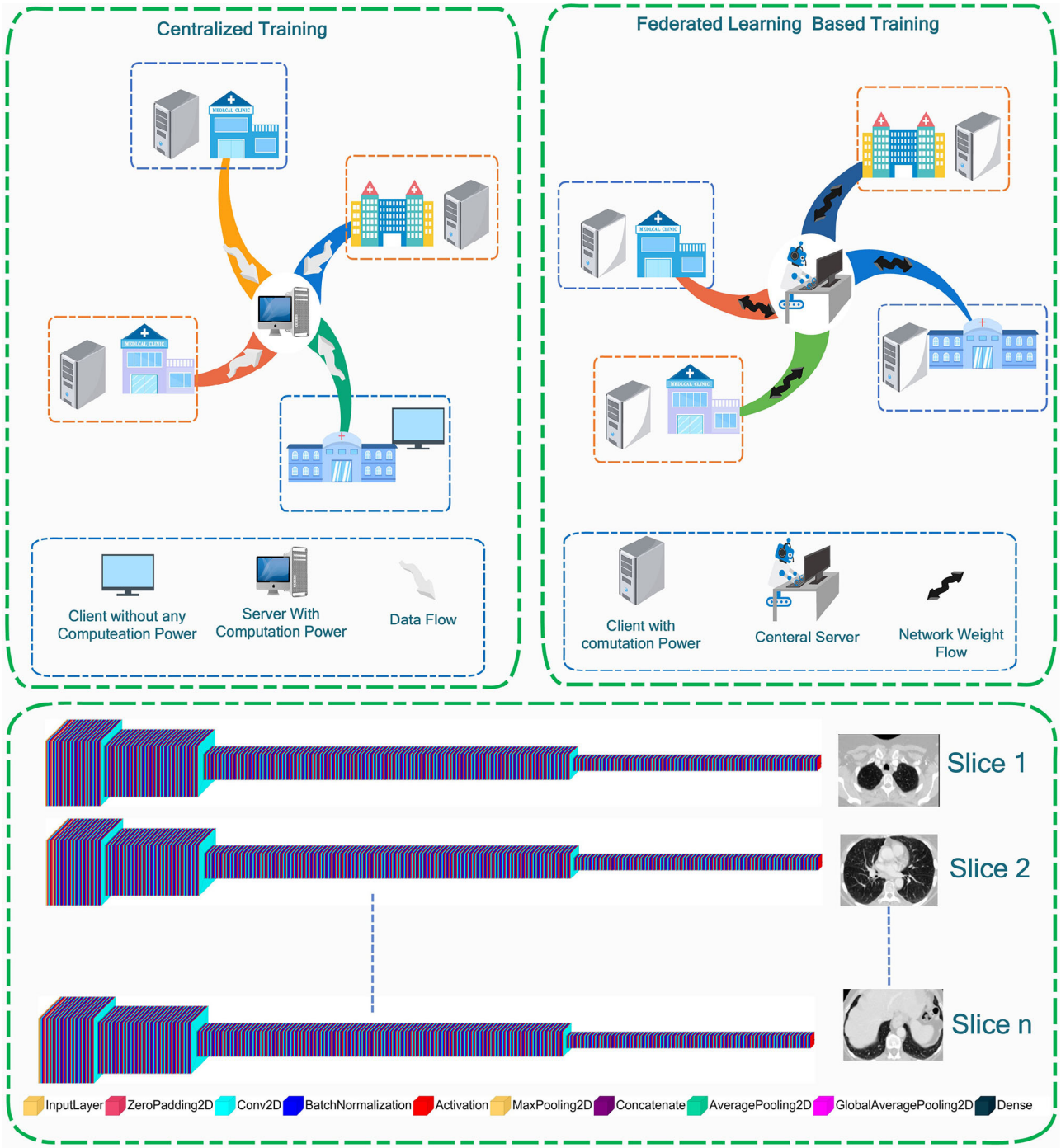

FIGURE 2 Centralized, federated learning and deep neural network schema were implemented in this study. In a centralized strategy, data are collected at multiple centers and pooled in a central location for training. In FL, data are collected at various centers, and a model is trained by distributing the training process across multiple centers without the need for data movement. We employed a combined 2D/3D model for the DL model core. Therefore, feature extraction was performed on 2D slices, and averaging was performed at the final layer during training to construct a 3D model for each scan. The DensNet model was used for 2D feature extraction.

function.[33,35] The adaptive quantile clipping function is defined as:

$$Clip\ (\theta, C_t) = \theta / max\left(1, \frac{\|\theta\|_2}{C_t}\right)$$

where θ is the model update from a decentralized data and C_t is the radius of the L_2 ball at iteration t.[33,35] This function bounds the L_2 norm of the model updates by projecting updates with a larger L_2 norm onto the L_2 ball of radius C_t.[33,35] The DP function adds Gaussian noise

24734209, 2024, 7, Downloaded from https://aapm.onlinelibrary.wiley.com/doi/10.1002/mp.16964 by Bibliotheque de l'Universite de Geneve, Division de l'information, Wiley Online Library on [10/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

to the model updates and is defined as:

$$DP\ (\theta) = \theta + \mathcal{N}\left(0, \frac{C_t^2}{\epsilon} I\right)$$

where θ is the model update from a decentralized data, $\mathcal{N}(0, \frac{C_t^2}{\epsilon} I)$ is a Gaussian noise vector with mean 0 and covariance matrix $\frac{C_t^2}{\epsilon} I$, C_t is the radius of the L_2 ball at iteration t, and ϵ is the privacy budget.[33,35]

In this approach, the global model updated by taking the sum of the clipped and differentially private model updates from all decentralized data sources and dividing them by the number of data sources.[33,35] The clipping function ensures that the L_2 norm of the model updates is bounded, while the DP function adds noise to protect the privacy of the data used to compute the model updates.[33,35] The hyperparameter C_t determines the radius of the L_2 ball and must be carefully chosen to balance the trade-off between model performance and robustness to update corruption.[33,35] The privacy budget ϵ determines the amount of noise added by the DP function and must be chosen such that the privacy of the data is adequately protected.[33,35]

2.8 | Overview of attacks and privacy analysis

In our study, we conducted a membership inference attack to assess the privacy risks associated with using the DL model.[20,21,54] In this scenario, the attacker attempts to determine whether specific cases of patient images were used for training the DL model.[20,21] In a membership inference attack, a binary classifier that predicts the given CT images was included in the training set for the target model.[20,21,54] Individual privacy is potentially compromised if the attacker models learn sufficient information regarding images.[20,21] We used the membership inference attack at intervals of every epoch to evaluate the privacy risks of DL models on medical images and identify potential measures for mitigating these risks.

2.9 | Model evaluation and statistical analysis

For model evaluation, metrics including precision, sensitivity, specificity, F1, accuracy, balanced accuracy, false negative rate (FNR), false positive rate (FPR), false discovery rate (FDR), negative predictive value (NPV), positive predictive value (PPV), and area under the receiver operating characteristic curve (AUC) were reported for the hold-out test sets (20% of each center data). In addition, models trained in two scenarios, including centralized and FL, were compared using the DeLong test for statistically significant differences. The significance level was considered at a level of 0.05.

TABLE 2 Summary of quantitative metrics across centralized and federated learning (rounded to two decimals). Metrics include precision, sensitivity (also known as recall or true positive rate), specificity, F1 score, accuracy, balanced accuracy, false negative rate (FNR), false discovery rate (FDR), false positive rate (FPR), negative predictive value (NPV), true positive rate (TPR), positive predictive value (PPV), and area under the curve (AUC) with mean and 95% confidence interval (95% CI). These metrics provide a comprehensive assessment of the model's predictive capabilities.

	Centralized	DPFL
Precision	0.74	0.73
Sensitivity	0.74	0.73
Specificity	0.77	0.77
F1	0.74	0.73
Accuracy	0.76	0.75
Balanced Accuracy	0.76	0.75
FNR	0.26	0.27
FDR	0.26	0.27
FPR	0.23	0.23
NPV	0.78	0.76
PPV	0.74	0.73
AUC (95% CI)	0.82 (0.79, 0.85)	0.81 (0.77, 0.84)

3 | RESULTS

Table 2 summarizes the statistics of the trained model in two different scenarios (rounded to two decimals). The centralized model had an accuracy of 0.76, while the DPFL model had 0.75. In terms of precision, the centralized model had a precision of 0.74, while the DPFL model had a precision of 0.73. Both the centralized and DPFL models had a specificity of 0.77. Finally, the centralized model had a sensitivity of 0.74, while the DPFL model had a sensitivity of 0.73.

The false positive rate was 0.23 for both the centralized and DPFL models, while the false discovery rate and false negative rate were 0.26 for the centralized model and 0.27 for the DPFL model. The negative predictive value was 0.78 for the centralized model and 0.76 for the DPFL model. A mean AUC of 0.82 and 0.81 was achieved with 95% confidence intervals (CI) of (0.79–0.85) for the centralized model and (0.77–0.84) for the DPFL model, respectively. The DeLong test did not prove statistically significant differences between the two models (p-value = 0.98). Figure 3 represents the AUC ROC curves for both models with 10000 bootstrapping and combined for comparing two different models.

In Figure 4, the performance of the DPFL model in relation to membership inference attacks during

24734209, 2024, 7, Downloaded from https://aapm.onlinelibrary.wiley.com/doi/10.1002/mp.16964 by Bibliotheque de l'Universite de Geneve, Division de l'information, Wiley Online Library on [10/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

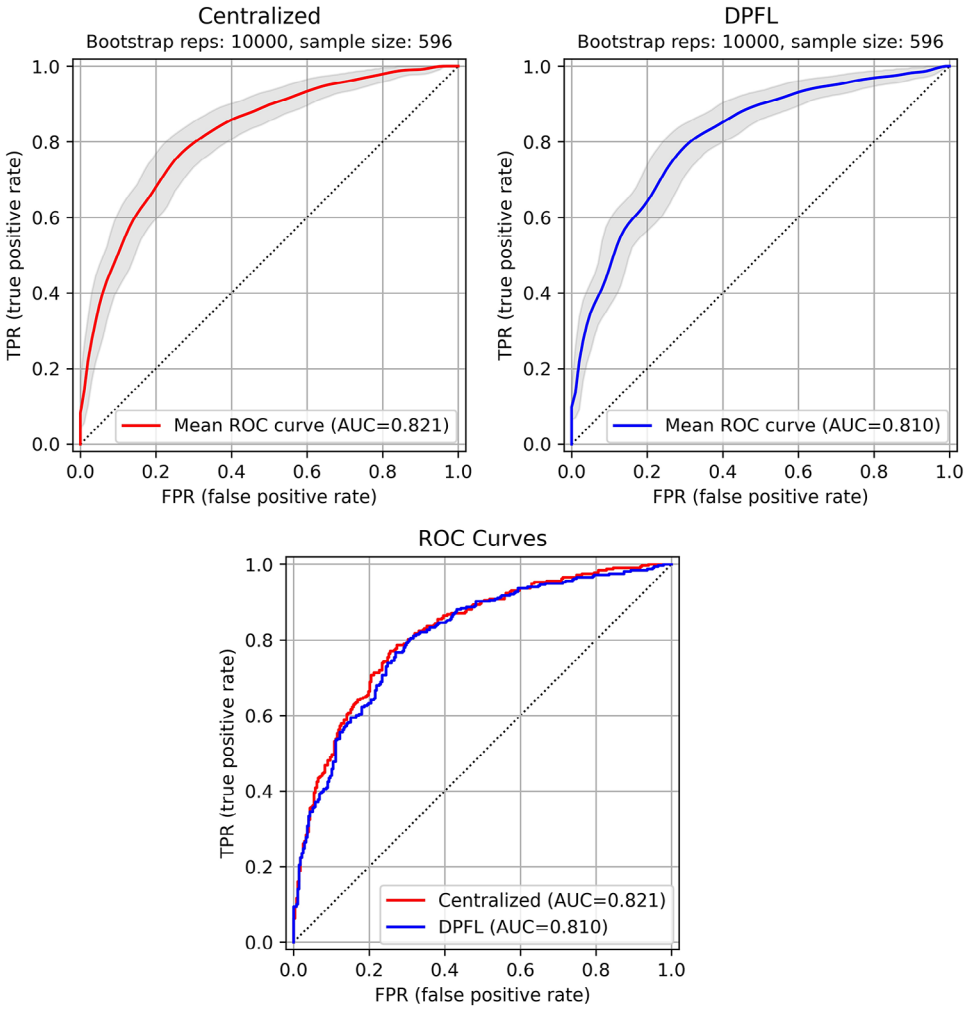

FIGURE 3 ROC curves of two different models, including centralized and differential privacy federated learning (DPFL), using 10000 bootstrapping and combined to compare two different models.

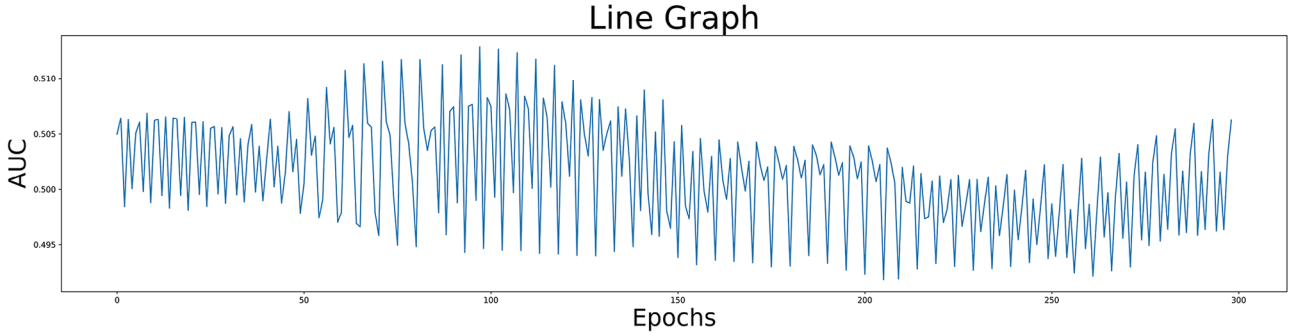

FIGURE 4 Line graph of different area under the curve (AUC) scores across different epochs for evaluating the resilience of the differential privacy federated learning (DPFL) model against membership inference attacks.

training is illustrated using the AUC metric. The AUC values for the membership inference attacks fluctuate between 0.49 and 0.518, with an average of 0.50 ± 0.003 and a 95% CI of 0.500 to 0.501. These results suggest that the membership inference attack was unable to successfully extract sensitive information regarding the patients during the training process. The DPFL model effectively protected patients' privacy with respect to membership inference of attack.

24734209, 2024, 7, Downloaded from https://aapm.onlinelibrary.wiley.com/doi/10.1002/mp.16964 by Bibliotheque de l'Universite de Geneve, Division de l'information, Wiley Online Library on [10/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

4 | DISCUSSION

FL has remained a decent solution to sensitive data transfer issues as it gives us the capacity to share the knowledge of the data without explicitly sharing the data itself.[20,21,55] Therefore, unlike centralized training, FL provides more comfortable options, such as access to multi-institutional data and computational infrastructure to build a generalizable DL model.[20,21] As a type of FL, DP-FL encompasses the advantage of keeping the data safe via the DP algorithm.[29] In the current study, we performed DPFL-based DL modeling for COVID-19 prognostication, which preserves patients' privacy and allows the use of multi-center model developments.

FL models face several threats (security and privacy)[20,21]; firstly, poisoning attack (model and data), in which an adversary attempts to manipulate the local updates of the model (e.g., by adding intentional errors or adversarial examples in the training set) and affects its overall performance in a negative way.[20,21,33,35,56] Secondly, models are prone to stealing attacks, which means that a copy of the global model along with its parameters is stolen as a whole, for example, by accessing the server.[20,21,57] Last but not least are membership inference attacks, in which the adversary sends queries to the model aiming to predict specific data points used during the training by the responses it receives.[20,21,58] Several methods were suggested to address the threats and attacks.[20,21] For instance, secure aggregation, in which model parameters are encrypted and then sent to the server, and computation (i.e., model aggregation) could be performed in encrypted or decrypted domains.[20,21,59] DP is another approach that adds some noise to the model updates or data.[20,21,60] Finally, there is the adversarial training method, which boosts the power of the FL model to defeat possible adversarial examples.[20,21,61]

Different studies have investigated the applicability of FL as a COVID-19 diagnostic and prognostic tool. For instance, Feki et al.[62] presented a deep FL for COVID-19 (COVID-19 and non-COVID-19) detection from 216 chest x-ray images. They evaluated the performance of their model, arguing it has comparative results to models trained by sharing patients' data.[62] Such results could encourage medical institutions to adopt collaborative processes with patient privacy preservation via FL approaches.[62] Another study was conducted by Liu et al.,[63] in which they started by raising concerns regarding the lack of sufficient data due to the fact that leakage and sharing of patients' health information are not allowed without permission,[63] and then proposed FL as a possible solution to the matter,[63] and compared the diagnostic accuracy of four popular models (MobileNet-v2, ResNet18, ResNet18, and COVID-Net) with and without FL frameworks.[63] They used the COVIDx dataset to train and evaluate a model for image classification.[63] Their dataset included 15 282 images, of which 13 703 and 1579 were allocated to training and testing, respectively.[63] They reported ResNet18 as the best-performing model with an accuracy of more than 0.90 in both FL and centralized scenarios.

Li et al.[64] designed a FedFocus for COVID-19 detection using 16 689 images, outperforming the baselines in model training, efficiency, accuracy, and stability. They used 8851, 6069, and 1769 normal, pneumonia, and COVID-19 samples, respectively, to build a model for a FL-powered detection tool.[64] Liang et al.[65] designed a FL framework utilizing 1 552 988 CT slices belonging to 4804 patients from 6 different centers, which outperformed the radiologist's assessment by diagnosing COVID-19 based on CT images alone with an AUC of 0.98. The electronic medical records (EMRs) and chest x-ray (EXAM) FL framework were presented for the clinical diagnosis of COVID-19 infection.[66] Using a total of 20 institutes from across the globe, it predicts the future oxygen requirements, mechanical ventilation, and death of symptomatic patients with COVID-19, using chest x-rays and clinical data.[66] With an average AUC greater than 0.92, EXAM successfully predicts outcomes at 24 and 72 h after patients initially present to the emergency room.[66]

Chowdhury et al.[67] developed DL models (Xception, ResNet50, DenseNet121, and InceptionV3) using 1823 images consisting of COVID-19 pneumonia, non-COVID-19 pneumonia, and healthy controls. They implemented FL through a pre-trained transfer learning model.[67] The performance metrics of the Xception model outperformed that of the rest with an accuracy of 0.9959.[67] However, the Inception V3 and DenseNet121 models resulted in metrics close to the Xception model with accuracies of 0.9917 and 0.9751, respectively.[67] Similarly, pre-trained DL models were utilized in a study by Florescue et al.[68] using 2230 axial chest CT images. The images were classified into three groups, including (i) COVID-19 cases, (ii) non-COVID-19 pathologies, for example, neoplastic diseases or non-COVID-19 viral pneumonia, and (iii) normal CT images.[68] The centralized and FL techniques were developed and achieved an accuracy of 0.79 and 0.7932 during the validation step, respectively.

Given the high variability of image acquisition parameters, particularly slice thickness across different centers, we opted for 2D feature extraction followed by an averaging method to construct a pseudo-3D model. This enabled to avoid the potential loss or distortion of information inherent in resizing for 3D processing and reduces computational load compared to zero padding to the largest dataset size (based on lung mask). In the current study, we implemented GDP-AQuCl, which theoretically addresses privacy and attack concerns.[20,21,33,35] We evaluated the model in a membership inference attack, which failed to provide sensitive information.[69,70] Nevertheless, there are some challenges to applying the DP mechanism as

24734209, 2024, 7, Downloaded from https://aapm.onlinelibrary.wiley.com/doi/10.1002/mp.16964 by Bibliotheque de l'Universite de Geneve, Division de l'information, Wiley Online Library on [10/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

well.[71–73] A major challenge is precisely determining a function's sensitivity, which is necessary to determine how much noise should be added.[71–73] Furthermore, adding noise to clean data can render statistical analysis less accurate, especially when a high magnitude of noise is required, that is, when the privacy budget is high.[71–73] The DPFL model, despite incorporating DP mechanisms, demonstrates performance on par with the centralized model. This result is achieved through careful calibration of DPFL parameters, ensuring an optimal balance between data privacy and model accuracy.

The current study suffers from some limitations, including the fact that the study was performed on one server using multiple GPUs.[33,35–37,74] Future studies should address the communication issues between different centers.[20,21,24,33,50] However, our study proved that model development is feasible for COVID-19 prognostication using chest CT images without sharing the data between centers while building global models, which achieved centralized model performances. In addition, more DL models and aggregation methods should be implemented for comprehensive evaluation. While our study evaluated the performance of the DL model on a hold-out test set, we recognize that validating the model on external datasets from different centers is essential to truly assess its generalizability and robustness. Future research should include such external validations to ensure the model's generalizability and robustness across different centers.

5 | CONCLUSION

This study evaluated the performance of deep FL, a privacy-preserving method, in predicting COVID-19 outcomes using chest CT images. Our results showed that the proposed model outperformed centralized models while operating on large and heterogeneous multi-institutional datasets. In addition, the model was resistant to inference attacks, ensuring the privacy of shared data during the training process. Overall, our results suggest that deep FL is a promising approach for predicting COVID-19 outcomes in a privacy-conscious manner. FL facilitates rapid and accurate data science and big-data collaboration between centers from across the globe, thus avoiding the need for sensitive data sharing.

AFFILIATIONS

[1]Division of Nuclear Medicine and Molecular Imaging, Geneva University Hospital, Geneva, Switzerland

[2]Research and Development Department, Med Fanavarn Plus Co, Karaj, Iran

[3]Department of Computer Science, University of Geneva, Geneva, Switzerland

[4]School of Medicine, Isfahan University of Medical Sciences, Isfahan, Iran

[5]Imaging Department, Qom University of Medical Sciences, Qom, Iran

[6]School of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran

[7]Department of Radiology Technology, Shahid Beheshti University of Medical Sciences, Tehran, Iran

[8]Department of Radiology, Shahid Beheshti Hospital, Qom University of Medical Sciences, Qom, Iran

[9]Department of Clinical Oncology, Royal Surrey County Hospital, Guildford, UK

[10]Radin Makian Azma Mehr Ltd., Radinmehr Veterinary Laboratory, Iran University of Medical Sciences, Gorgan, Iran

[11]Department of Medical Physics, Faculty of Medical Sciences, Tarbiat Modares University, Tehran, Iran

[12]Clinical Research Development Unit (CRDU), Sayad Shirazi Hospital, Golestan University of Medical Sciences, Gorgan, Iran

[13]Medical Imaging Research Center, Department of Radiology, Shiraz University of Medical Sciences, Shiraz, Iran

[14]Department of Radiology, School of Medicine, Advanced Diagnostic and Interventional Radiology Research Center (ADIR), Imam Khomeini Hospital, Tehran University of Medical Sciences, Tehran, Iran

[15]Digestive Diseases Research Center, Digestive Diseases Research Institute, Tehran University of Medical Sciences, Tehran, Iran

[16]Department of Neurosurgery, Faculty of Medical Sciences, Iran University of Medical Sciences, Tehran, Iran

[17]Department of Immunology, School of Medicine, Fasa University of Medical Sciences, Fasa, Iran

[18]Department of Radiology, Arak University of Medical Sciences, Arak, Iran

[19]Clinical research development center, Qom University of Medical Sciences, Qom, Iran

[20]Department of radiation oncology, Shohada-e Tajrish Hospital, Shahid Beheshti University of Medical Sciences, Tehran, Iran

[21]Infectious Diseases Research Center, Golestan University of Medical Sciences, Gorgan, Iran

[22]Department of Infectious Disease, School of Medicine, Fasa University of Medical Sciences, Fasa, Iran

[23]Department of Radiology, Shariati Hospital, Tehran University of Medical Sciences, Tehran, Iran

[24]Rajaie Cardiovascular, Medical & Research Center, Iran University of Medical Science, Tehran, Iran

[25]Department of Computer Science, University of British Columbia, Vancouver, British Columbia, Canada

[26]Department of Public Health Sciences, College of Medicine, Pennsylvania State University, Hershey, Pennsylvania, USA

[27]Department of Nuclear Medicine and Molecular Imaging, University of Groningen, University Medical Center Groningen, Groningen, Netherlands

[28]Department of Nuclear Medicine, University of Southern Denmark, Odense, Denmark

[29]University Research and Innovation Center, Óbuda University, Budapest, Hungary

ACKNOWLEDGMENTS

The Swiss National Science Foundation supported this work under grant SNSF 320030_176052.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

REFERENCES

1. Collaborators C-EM. Estimating excess mortality due to the COVID-19 pandemic: a systematic analysis of COVID-19-related mortality, 2020–21. *Lancet*. 2022;399(10334):1513-1536. [Published online ahead of print 20220310].
2. Wynants L, Van Calster B, Collins GS, et al. Prediction models for diagnosis and prognosis of covid-19: systematic review and

24734209, 2024, 7, Downloaded from https://aapm.onlinelibrary.wiley.com/doi/10.1002/mp.16964 by Bibliotheque de l'Universite de Geneve, Division de l'information, Wiley Online Library on [10/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

critical appraisal. *BMJ*. 2020;369:m1328. [Published online ahead of print 20200407].
3. Li X, Li T, Wang H. Treatment and prognosis of COVID-19: current scenario and prospects. *Exp Ther Med*. 2021;21(1):1-1.
4. Gronholm PC, Nosé M, van Brakel WH, et al. Reducing stigma and discrimination associated with COVID-19: early stage pandemic rapid review and practical recommendations. *Epidemiol Psychiatr Sci*. 2021;30:e15. [Published online ahead of print 20210128].
5. Francone M, Iafrate F, Masci GM, et al. Chest CT score in COVID-19 patients: correlation with disease severity and short-term prognosis. *Eur Radiol*. 2020;30(12):6808-6817. [Published online ahead of print 20200704].
6. Prokop M, Everdingen Wv, Vellinga TvR, et al. CO-RADS: a categorical CT assessment scheme for patients suspected of having COVID-19—definition and evaluation. *Radiology*. 2020;296(2):E97-E104.
7. Filchakova O, Dossym D, Ilyas A, Kuanysheva T, Abdizhamil A, Bukasov R. Review of COVID-19 testing and diagnostic methods. *Talanta*. 2022;244:123409. [Published online ahead of print 20220331].
8. Ye Z, Zhang Y, Wang Y, Huang Z, Song B. Chest CT manifestations of new coronavirus disease 2019 (COVID-19): a pictorial review. *Eur Radiol*. 2020;30(8):4381-4389. [Published online ahead of print 20200319].
9. Rizzetto F, Berta L, Zorzi G, et al. Diagnostic performance in differentiating COVID-19 from other viral pneumonias on CT imaging: multi-reader analysis compared with an artificial intelligence-based model. *Tomography*. 2022;8(6):2815-2827. [Published online ahead of print 20221125].
10. Pu J, Leader JK, Bandos A, et al. Automated quantification of COVID-19 severity and progression using chest CT images. *Eur Radiol*. 2021;31(1):436-446. [Published online ahead of print 20200813].
11. Elmokadem AH, Mounir AM, Ramadan ZA, Elsedeiq M, Saleh GA. Comparison of chest CT severity scoring systems for COVID-19. *Eur Radiol*. 2022;32(5):3501-3512. [Published online ahead of print 20220115].
12. Carbonell G, Del Valle DM, Gonzalez-Kozlova E, et al. Quantitative chest CT combined with plasma cytokines predict outcomes in COVID-19 patients. *medRxiv*. 2021. doi:10.1101/2021.10.11.21264709
13. Li K, Wu J, Wu F, et al. The clinical and chest CT features associated with severe and critical COVID-19 pneumonia. *Invest Radiol*. 2020;55(6):327-331. [Published online ahead of print 2020/03/03].
14. Shiri I, Salimi Y, Pakbin M, et al. COVID-19 prognostic modeling using CT radiomic features and machine learning algorithms: analysis of a multi-institutional dataset of 14,339 patients. *Comput Biol Med*. 2022;145:105467. [Published online ahead of print 20220329].
15. Shiri I, Mostafaei S, Haddadi Avval A, et al. High-dimensional multinomial multiclass severity scoring of COVID-19 pneumonia using CT radiomics features and machine learning algorithms. *Sci Rep*. 2022;12(1):14817. [Published online ahead of print 20220901].
16. Moradi Khaniabadi P, Bouchareb Y, Al-Dhuhli H, et al. Two-step machine learning to diagnose and predict involvement of lungs in COVID-19 and pneumonia using CT radiomics. *Comput Biol Med*. 2022;150:106165. [Published online ahead of print 20221005].
17. Shiri I, Salimi Y, Saberi A, et al. Differentiation of COVID-19 pneumonia from other lung diseases using CT radiomic features and machine learning: a large multicentric cohort study. *Int J Imaging Syst Technol*. 2024;34(2):e23028.
18. Gong K, Wu D, Arru CD, et al. A multi-center study of COVID-19 patient prognosis using deep learning-based CT image analysis and electronic health records. *Eur J Radiol*. 2021;139:109583. [Published online ahead of print 20210205].
19. Wang S, Zha Y, Li W, et al. A fully automatic deep learning system for COVID-19 diagnostic and prognostic analysis. *Eur Respir J*. 2020;56(2):2000775. [Published online ahead of print 20200806].
20. Rahman KJ, Ahmed F, Akhter N, et al. Challenges, applications and design aspects of federated learning: a survey. *IEEE Access*. 2021;9:124682-124700.
21. Prayitno, Shyu C-R, Putra KT, et al. A systematic review of federated learning in the healthcare area: from the perspective of data properties and applications. *Appl Sci*. 2021;11(23):11191.
22. Kaissis G, Ziller A, Passerat-Palmbach J, et al. End-to-end privacy preserving deep learning on multi-institutional medical imaging. *Nat Mach Intell*. 2021;3(6):473-484.
23. Wen H, Zhao Q, Lin Z, Xuan D, Shroff N. A study of the privacy of covid-19 contact tracing apps. Paper presented at: Security and Privacy in Communication Networks: 16th EAI International Conference, SecureComm 2020, Washington, DC, USA, October 21–23, 2020, Proceedings, Part I 162020.
24. Jin H, Luo Y, Li P, Mathew J. A review of secure and privacy-preserving medical data sharing. *IEEE Access*. 2019;7:61656-61669.
25. Kaissis GA, Makowski MR, Rückert D, Braren RF. Secure, privacy-preserving and federated machine learning in medical imaging. *Nat Mach Intell*. 2020;2(6):305-311.
26. Shi M, Jiang R, Hu X, Shang J. A privacy protection method for health care big data management based on risk access control. *Health Care Manag Sci*. 2020;23(3):427-442. [Published online ahead of print 20190723].
27. Nelson GS. Practical implications of sharing data: a primer on data privacy, anonymization, and de-identification. Paper presented at: SAS global forum proceedings. 2015.
28. Chen X, Hu C-J. Adaptive medical image encryption algorithm based on multiple chaotic mapping. *Saudi J Biol Sci*. 2017;24(8):1821-1827.
29. Wei K, Li J, Ding M, et al. Federated learning with differential privacy: algorithms and performance analysis. *IEEE Trans Inf Forensics Secur*. 2020;15:3454-3469.
30. Liu R, Gupta S, Patel P. The application of the principles of responsible AI on social media marketing for digital health. *Inf Syst Front*. 2021:1-25.
31. Kumar P, Dwivedi YK, Anand A. Responsible artificial intelligence (AI) for value formation and market performance in healthcare: the mediating role of patient's cognitive engagement. *Inf Syst Front*. 2021:1-24.
32. Ho TT, Tran KD, Huang Y. FedSGDCOVID: federated SGD COVID-19 detection under local differential privacy using chest x-ray images and symptom information . *Sensors (Basel)*. 2022;22(10):3728. [Published online ahead of print 20220513].
33. Shiri I, Razeghi B, Vafaei Sadr A, et al. Multi-institutional PET/CT image segmentation using federated deep transformer learning. *Comput Methods Programs Biomed*. 2023;240:107706. [Published online ahead of print 20230712].
34. Shiri I, Sadr AV, Sanaat A, Ferdowsi S, Arabi H, Zaidi H. Federated learning-based deep learning model for PET attenuation and scatter correction: a multi-center study. Paper presented at: 2021 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC). 2021.
35. Shiri I, Salimi Y, Maghsudi M, et al. Differential privacy preserved federated transfer learning for multi-institutional (68)Ga-PET image artefact detection and disentanglement. *Eur J Nucl Med Mol Imaging*. 2023;51(1):40-53. [Published online ahead of print 20230908].
36. Shiri I, Vafaei Sadr A, Akhavan A, et al. Decentralized collaborative multi-institutional PET attenuation and scatter correction using federated deep learning. *Eur J Nucl Med Mol Imaging*. 2023;50(4):1034-1050. [Published online ahead of print 20221212].
37. Shiri I, Vafaei Sadr A, Amini M, et al. Decentralized distributed multi-institutional PET image segmentation using a federated

24734209, 2024, 7, Downloaded from https://aapm.onlinelibrary.wiley.com/doi/10.1002/mp.16964 by Bibliotheque de l'Universite de Geneve, Division de l'information, Wiley Online Library on [10/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

deep learning framework. *Clin Nucl Med*. 2022;47(7):606-617. [Published online ahead of print 20220420].
38. Shiri I, Razeghi B, Ferdowsi S, et al. PRIMIS: privacy-preserving medical image sharing via deep sparsifying transform learning with obfuscation. *J Biomed Inform*. 2024;150:104583. doi:10.1016/j.jbi.2024.104583:104583
39. Abdul Salam M, Taha S, Ramadan M. COVID-19 detection using federated machine learning. *PLoS One*. 2021;16(6):e0252573. [Published online ahead of print 20210608].
40. Ning W, Lei S, Yang J, et al. Open resource of clinical data from patients with pneumonia for the prediction of COVID-19 outcomes via deep learning. *Nat Biomed Eng*. 2020;4(12):1197-1207. [Published online ahead of print 20201118].
41. Radpour A, Bahrami-Motlagh H, Taaghi MT, et al. COVID-19 evaluation by low-dose high resolution CT scans protocol. *Acad Radiol*. 2020;27(6):901-901.
42. Shiri I, Arabi H, Salimi Y, et al. COLI-Net: deep learning-assisted fully automated COVID-19 lung and infection pneumonia lesion detection and segmentation from chest computed tomography images. *Int J Imaging Syst Technol*. 2022;32(1):12-25.
43. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. Paper presented at: Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.
44. Li L, Fan Y, Tse M, Lin K-Y. A review of applications in federated learning. *Comput Ind Eng*. 2020;149:106854.
45. Rieke N, Hancox J, Li W, et al. The future of digital health with federated learning. *NPJ Digit Med*. 2020;3(1):119.
46. AbdulRahman S, Tout H, Ould-Slimane H, Mourad A, Talhi C, Guizani M. A survey on federated learning: the journey from centralized to distributed on-site learning and beyond. *IEEE Internet Things J*. 2020;8(7):5476-5497.
47. Rahman SA, Tout H, Talhi C, Mourad A. Internet of things intrusion detection: centralized, on-device, or federated learning? *IEEE Network*. 2020;34(6):310-317.
48. Pillutla K, Kakade SM, Harchaoui Z. Robust aggregation for federated learning. *IEEE Transactions on Signal Processing*. 2022;70:1142-1154.
49. Li C, Li G, Varshney PK. Communication-efficient federated learning based on compressed sensing. *IEEE Internet Things J*. 2021;8(20):15531-15541.
50. Li Q, Wen Z, Wu Z, et al. A survey on federated learning systems: vision, hype and reality for data privacy and protection. *IEEE Trans Knowl Data Eng*. 2021.
51. Andrew G, Thakkar O, McMahan B, Ramaswamy S. Differentially private learning with adaptive clipping. *Adv Neural Inf Process Syst*. 2021;34:17455-17466.
52. Chen W-N, Choo CAC, Kairouz P, Suresh AT. The fundamental price of secure aggregation in differentially private federated learning. Paper presented at: International Conference on Machine Learning. 2022.
53. Liu Z, Guo J, Yang W, Fan J, Lam K-Y, Zhao J. Privacy-preserving aggregation in federated learning: a survey. *IEEE Trans Big Data*. 2022.
54. Hu H, Salcic Z, Sun L, Dobbie G, Yu PS, Zhang X. Membership inference attacks on machine learning: a survey. *ACM Comput Surv (CSUR)*. 2022;54(11s):1-37.
55. Xu J, Glicksberg BS, Su C, Walker P, Bian J, Wang F. Federated learning for healthcare informatics. *J Healthc Inform Res*. 2021;5:1-19.
56. Xie C, Huang K, Chen P-Y, Li B. Dba: distributed backdoor attacks against federated learning. Paper presented at: International conference on learning representations. 2020.
57. Jere MS, Farnan T, Koushanfar F. A taxonomy of attacks on federated learning. *IEEE Secur Priv*. 2020;19(2):20-28.
58. Gu Y, Bai Y, Xu S. CS-MIA: membership inference attack based on prediction confidence series in federated learning. *J Inf Secur Appl*. 2022;67:103201.
59. Bonawitz K, Ivanov V, Kreuter B, et al. Practical secure aggregation for federated learning on user-held data. *arXiv preprint arXiv:161104482*. 2016.
60. Lyu L, Yu H, Yang Q. Threats to federated learning: a survey. *arXiv preprint arXiv:200302133*. 2020.
61. Bhagoji AN, Chakraborty S, Mittal P, Calo S. Analyzing federated learning through an adversarial lens. Paper presented at: International Conference on Machine Learning. 2019.
62. Feki I, Ammar S, Kessentini Y, Muhammad K. Federated learning for COVID-19 screening from Chest X-ray images. *Applied soft computing*. 2021;106:107330. [Published online ahead of print 20210320].
63. Liu B, Yan B, Zhou Y, Yang Y, Zhang Y. Experiments of federated learning for covid-19 chest x-ray images. *arXiv preprint arXiv:200705592*. 2020.
64. Li Z, Xu X, Cao X, et al. Integrated CNN and Federated Learning for COVID-19 Detection on Chest X-Ray Images. *IEEE/ACM Trans Comput Biol Bioinform*. 2022;1-11. [Published online ahead of print 20220620].
65. Liang H, Guo Y, Chen X, et al. Artificial intelligence for stepwise diagnosis and monitoring of COVID-19. *Eur Radiol*. 2022;32(4):2235-2245. [Published online ahead of print 20220106].
66. Dayan I, Roth HR, Zhong A, et al. Federated learning for predicting clinical outcomes in patients with COVID-19. *Nat Med*. 2021;27(10):1735-1743. [Published online ahead of print 20210915].
67. Chowdhury D, Banerjee S, Sannigrahi M, et al. Federated learning based Covid-19 detection. *Expert Syst*. 2023:e13173. doi:10.1111/exsy.13173:e13173
68. Florescu LM, Streba CT, Şerbănescu MS, et al. Federated learning approach with pre-trained deep learning models for COVID-19 detection from unsegmented CT images. *Life (Basel)*. 2022;12(7):958. [Published online ahead of print 20220626].
69. Mothukuri V, Parizi RM, Pouriyeh S, Huang Y, Dehghantanha A, Srivastava G. A survey on security and privacy of federated learning. *Future Gener Comput Syst*. 2021;115:619-640.
70. Dwork C, Roth A. The algorithmic foundations of differential privacy. *Found Trends Theor Comput Sci*. 2014;9(3–4):211-407.
71. Machanavajjhala A, He X, Hay M. Differential privacy in the wild: a tutorial on current practices & open challenges. Paper presented at: Proceedings of the 2017 ACM International Conference on Management of Data. 2017.
72. Sarwate AD, Chaudhuri K. Signal processing and machine learning with differential privacy: algorithms and challenges for continuous data. *IEEE Signal Process Mag*. 2013;30(5):86-94.
73. Garfinkel SL, Abowd JM, Powazek S. Issues encountered deploying differential privacy. Paper presented at: Proceedings of the 2018 Workshop on Privacy in the Electronic Society. 2018.
74. Shiri I, Amini M, Salimi Y, et al. Multi-institutional PET/CT image segmentation using a decentralized federated deep transformer learning algorithm. *J Nucl Med*. 2022;63(2):3218.

How to cite this article: Shiri I, Salimi Y, Sirjani N, et al. Differential privacy preserved federated learning for prognostic modeling in COVID-19 patients using large multi-institutional chest CT dataset. *Med Phys*. 2024;51:4736–4747. https://doi.org/10.1002/mp.16964

24734209, 2024, 7, Downloaded from https://aapm.onlinelibrary.wiley.com/doi/10.1002/mp.16964 by Bibliotheque de l'Universite de Geneve, Division de l'information, Wiley Online Library on [10/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License