Image Anal Stereol 2024;43:1-8 doi: 105566/ias.3131
Original Research Paper

IMPROVING THE LEARNING PERFORMANCE OF CLIENT'S LOCAL DISTRIBUTION IN CYCLIC FEDERATED LEARNING

LI KANG[1,2], BIN LUO[1,2] AND JIANJUN HUANG[1,2]✉

[1]Guangdong Key Laboratory of Intelligent Information Processing, Shenzhen, China; [2]College of Electronics and Information Engineering, Shenzhen University, Shenzhen, Guangdong, China
e-mail: kangli@szu.edu.cn; 2210433025@email.szu.edu.cn; huangjin@szu.edu.cn
(Received January 9, 2024; revised February 18, 2024; accepted February 18, 2024)

ABSTRACT

Cyclic federated learning based on distribution information sharing and knowledge distillation (CFL_DS_KD) aims to address the challenges of non-iid data distribution and reduce communication requirements. However, when client data is extremely heterogeneous and scarce, it becomes challenging for clients to fully learn the distribution of local data using GANs, thereby affecting the overall model performance. To overcome this limitation, we propose a transfer learning approach where clients first pretrain their generators on a source domain and then fine-tune them on their local datasets. Our results on the classification of Alzheimer's disease demonstrate that this method effectively improves client distribution learning performance and enhances the overall model performance.

Keywords: federated learning, medical image processing, transfer learning.

INTRODUCTION

Deep learning has found widespread applications in intelligent healthcare (Miotto *et al.*, 2017), including disease prediction, diagnosis, treatment, and prognosis. However, training effective deep learning models often requires large centralized datasets, which pose a significant challenge in areas where data privacy is crucial. Due to the sensitive nature of medical data, patient data from different hospitals cannot be exchanged or centrally stored. As a result, traditional deep learning models lack publicly shared medical datasets for training. Federated learning has emerged as a promising solution to address the privacy concerns associated with data. By enabling distributed learning, federated learning allows multiple organizations to collaboratively train a global model while preserving data privacy (Yang *et al.*, 2019). However, due to the non-iid nature of datasets from different institutions, local models trained on individual datasets may overfit, leading to poor generalization of the global model. Distribution sharing among clients is a promising approach to address the non-iid problem. However, if the local client's data is scarce and extremely heterogeneous, the ability of the local client to learn the local distribution will be compromised, resulting in poor quality of the shared distribution information. In this work, we propose to utilize transfer learning to improve the learning performance of client's local distribution. Specifically, we first utilize a GAN model to learn the data distribution in the source domain and then fine-tune it in the target domain. This process aims to enhance the ability of the GAN model to learn the local data distribution. Subsequently, we apply the improved GAN model in the cyclic federated learning method based on the distribution of information sharing and knowledge distillation (CFL_DS_KD) (Yu *et al.*, 2022) for classification of Alzheimer's disease.

RELATED WORKS

The non-iid challenge in federated learning

Federated learning (FL) (McMahan *et al.*, 2017) involves training statistical models over remote de-vices or siloed data centers, such as mobile phones or hospitals, while keeping data localized. A major challenge in FL is that the data across clients is not identically and independently distributed (non-iid). In response to non-iid problems, existing research has mainly solved the problems at the algorithm and data levels. The algorithm-level solutions mainly include objective function modification and solution mode optimization. Objective function modification in-volves adding regularization terms on the client side. A trade-off has been achieved between optimizing local models and reducing the differences between local models and global models to solve the non-independent homogeneous distribution of data at each node. For example, FedProx (Li *et al.*, 2020) has been proposed to corrects the client-side drift that occurs in FedAvg (McMahan *et al.*, 2017) by restricting

the Euclidean distances between local models and global models as proximal terms. This means that the local updates do not excessively deviate from the global models, which alleviates any inconsistencies in the client-side data and improves the stability of global model convergence. FedCurv (Shoham *et al.*, 2019) uses Fisher information from global models obtained during the previous rounds of training to weight the distances, which can reduce excessive errors in the model parameters. SCAFFOLD (Karimireddy *et al.*, 2020) has been proposed to improve the FedProx by adding a control variable on the client side. This control variable can take either the gradient norm of global models on local datasets or the Euclidean distances between local and global models, thus preventing local models from deviating from the globally correct training direction. These methods can improve the performance of federated learning for model learning on non-iid datasets to some extent, but the degree of improvement is limited by the consistency of the client-side data sampling.

In solution optimization, the good performance of federated learning models is mainly achieved by improving the server-side aggregation method. FedAvg determines client aggregation weights based on the size of clients 'datasets. However, in non-iid scenarios, this aggregation method leads to a significant decrease in the performance of the global model. For this reason, most scholars have aimed to seek better aggregation method.

In ABAvg (Xiao *et al.*, 2021), the server-side tests the accuracy of temporary models on validation datasets to obtain the accuracy of the models on the client side and then normalizes them before aggregating all parameters. FedMA (Wang *et al.*, 2020) uses Bayesian nonparametric methods to match and average weights in a hierarchical manner. FedAvgM (Tsu *et al.*, 2019) applies momentum when updating global models on a server. FedNova (Wang *et al.*, 2020) normalizes local updates before averaging. However, these methods have limited success in improving the performance of global models (Karimireddy *et al.*, 2020), so some scholars have proposed approaches that evade this problem, such as personalized federated learning, multitask federated learning and federated meta-learning, which can also improve the performance of federated learning on non-iid data to some extent.

Transfer learning for medical data

Transfer learning (TL) stems from cognitive research, which uses the idea, that knowledge is transferred across related tasks to improve performances on a new task. The formal definition of TL is defined by Pan and Yang with the notions of domains and tasks. A domain consists of a feature space $\mathcal{X}$ and marginal probability distribution $P(X)$, where $X = \{x_1, \dots, x_n\} \in \mathcal{X}$. Given a specific domain denoted by $D = \{\mathcal{X}, P(X)\}$, a task is denoted by $T = \{\mathcal{Y}, f(\cdot)\}$ where $\mathcal{Y}$ is a label space and $f(\cdot)$ is an objective predictive function. Given a source domain $\boldsymbol{D_S}$ and learning task $\boldsymbol{T_S}$, a target domain $\boldsymbol{T_S}$ and learning task $\boldsymbol{T_T}$, transfer learning aims to improve the learning of the target predictive function $f_T(\cdot)$ in $\boldsymbol{D_T}$ by using the knowledge in $\boldsymbol{D_S}$ and $\boldsymbol{D_S}$ (Pan. *et al*., 2020).

There have been lots of studies applying transfer learning to medical image processing. Swati *et al.* use pre-trained deep CNN model and propose a block-wise fine-tuning strategy based on transfer learning which is evaluated on T1-weighted contrast-enhanced magnetic resonance images (CE-MRI) benchmark dataset. Experimental results show that their proposed method outperforms state-of-the-art classification on the CE-MRI dataset. da Nóbrega *et al.* trained several CNN (e.g.VGG16, MobileNet, ResNet50, DenseNet169, etc.) on the ImageNet dataset, converted them into feature extractors and applied on the LIDC/IDRI nodule images. Hassan *et al.* proposed an efficient and accurate approach for medical image modality classification which is developed using transfer learning concept with pretrained ResNet50 Deep learning model for optimized features extraction followed by linear discriminant analysis classification (TLRN-LDA). Gessert *et al.* demonstrate that convolutional neural networks and transfer learning can be used to identify cancer tissue with confocal laser microscopy and show that there is no generally optimal transfer learning strategy and model as well as task-specific engineering is required.

METHOD

Learning client data distribution through transfer learning

To acquire knowledge about the distribution of hospital data, deep learning-based generator models are commonly employed. Generators are highly effective for data augmentation as they can learn the distribution information of data and generate data that aligns with the actual distribution. Generative adversarial networks (GANs) are a prevalent class of deep neural network generators known for their re-markable capabilities in image enhancement and image-to-image conversion. In our study, we utilize GANs as data generators to capture the data distribution information from local clients. However, considering the limited availability and heterogeneity of local data, allowing GANs to directly train on local data may pose challenges in fully capturing the underlying distribution. Therefore, we propose using

transfer learning to enhance the learning of distribution information by local clients. Specifically, as shown in Fig. 1, we first allow local generators to learn distribution knowledge in the source domain and then fine-tune them using data from local clients.

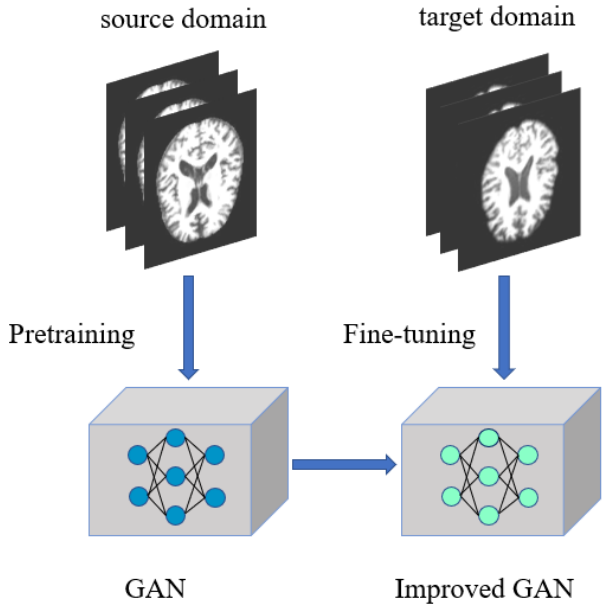


Fig. 1. *Improving the performance of GAN in learning client distributions through transfer learning.*

Model pre-training serves to minimize internal dimensions and implicitly influences the model's induction bias. In classical supervised learning, models often possess a strong inductive bias, such as the local connectivity assumptions in convolutional neural networks (CNNs) and recurrent neural networks (RNNs). Pre-training provides an inductive bias for downstream tasks, which often have limited labeled samples, enabling the pre-trained model parameters (with hundreds of millions of samples) to generalize well when fine-tuned with a small amount of data. The core idea of our method involves pre-training the GAN in the source domain to extract features and initialize the GAN network parameters. Subsequently, fine-tuning is performed in the target domain. Transfer learning, in this context, aims to enhance the model's performance by identifying differences between datasets and leveraging transferable knowledge. Generative adversarial networks, designed to generate similar data by approximating the feature distribution of the target samples, typically require a sufficient number of target samples. When the target sample size is small, GANs often face mode collapse issues. However, transfer learning can alleviate this problem in GANs and reduce the stringent requirement of similarity between the source and target domain data.

In general, by leveraging transfer learning, GANs can effectively learn the distribution of local datasets, thereby enhancing the quality of generated medical images by local GANs. These improved models can then be applied in the context of cyclic federated learning.

Transfer learning based CFL_DS_KD

Once we have acquired a well-trained GAN model that effectively captures the data distribution of the client through transfer learning, we can proceed to integrate it into the cyclic federated learning method, which relies on the sharing of distribution information and knowledge distillation. The specific steps are as follows: Let C represent the total number of clients participating in the federated learning task. Let $D_c = \{x_i | i = 1,2 \dots, N_c\}$ be the local dataset of client c (where $c = 1,2, \dots, C$) and $N_c = |D_c|$ be the number of samples in the local dataset. Initially, the client c trains a generator through transfer learning which reflects the distribution information G_c of local datasets D_c. Thus, C clients are trained to obtain C generator models. Then, client c transmit its generator G_c to the client $c + 1$, forming a ring-shaped communication link when $c = C$ let $c + 1 = c$. Then, the generator G_c from the client c can generate N'_{c+1} virtually shared data points, i.e., $D'_{c+1} = \{x_l | x_l = G(z_l), l = 1,2, \dots, N'_{c+1}\}$. The distribution information sharing process is schematically illustrated in Fig. 2. Additionally, client c transmits local pretrained model ω_c to the client $c + 1$.

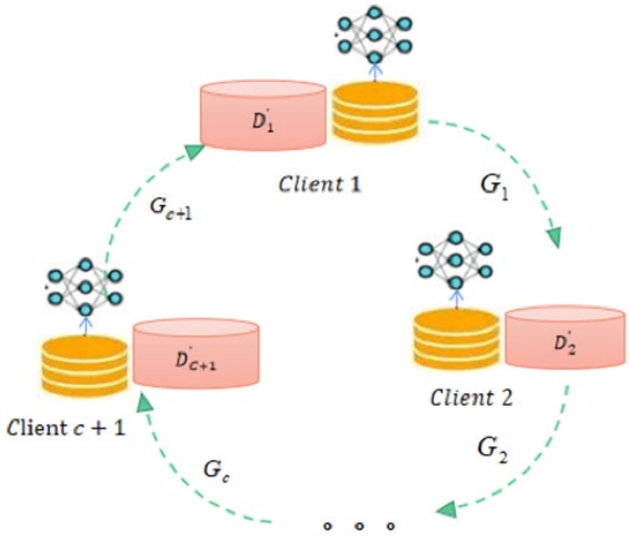


Fig. 2. *The process of client distribution sharing and virtual dataset generation in cyclic federated learning.*

Then, the $c + 1$ client employs knowledge distillation, utilizing ω_c as a teacher model to guide the training of ω_{c+1} on the virtual dataset N'_{c+1}, as shown in Fig. 3. After the process of knowledge distillation, the updated

model of client $c+1$ continues to train on the local dataset D_{c+1}.

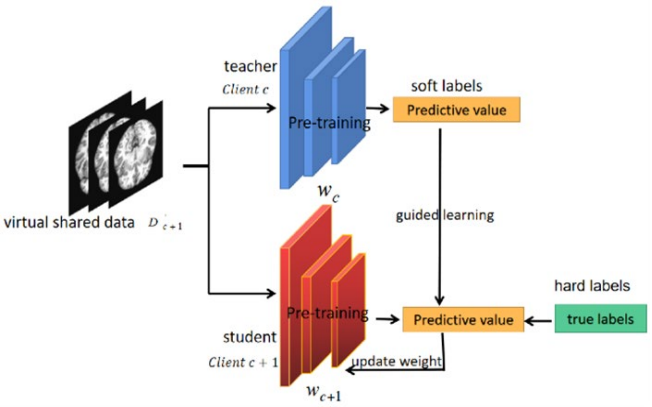


Fig. 3. *The process of knowledge distillation on virtual shared data generated by GAN.*

The training goal of the cyclic federated learning method based on the distribution of information sharing and knowledge distillation was the minimization of the total loss function (1).

$$l(\omega_1, \omega_2, \dots, \omega_c) = \sum_{c=1}^{C} L_{c+1}(\omega_{c+1}) + \lambda \sum_{c=1}^{C} R_{c+1}(\omega_{c+1}, \omega_c) \quad (1)$$

$$L_{c+1}(\omega_{c+1}) = \sum_{x \in D_{c+1}} l_{c+1}(x; \omega_{c+1}) \quad (2)$$

$$R_{c+1}(\omega_{c+1}, \omega_c) = \alpha L_{soft}(\omega_{c+1}, \omega_c) + \beta l_{hard}(\omega_{c+1}) \quad (3)$$

$$L_{soft}(\omega_{c+1}, \omega_c) = \sum_{x \in D'_{c+1}} l_{soft}(x; \omega_{c+1}, \omega_c) \quad (4)$$

$$L_{hard}(\omega_{c+1}) = \sum_{x \in D'_{c+1}} l_{hard}(x; \omega_{c+1}) \quad (5)$$

In (1), R_{c+1} represents the loss of client $c+1$ during training on the virtual dataset D'_{c+1} using the model ω_c of client c. As shown in equation (3), this loss includes both the soft loss during the knowledge distillation process and the hard loss of the student model. L_{c+1} represents the loss of client $c+1$ during training on the local dataset. Equations (4) and (5) describe the optimization process which indicates that $\omega_{c+1}^{(k-1)}$ is first optimized through training on the virtual dataset to obtain the updated model $\mu_{c+1}^{(k)}$, and then further updated on the local dataset to obtain the final model $\omega_{c+1}^{(k)}$. Then, the updated model $\omega_{c+1}^{(k)}$ transmits to next client.

$$\mu_{c+1}^{(k)} = \omega_{c+1}^{(k-1)} - \alpha_k \nabla R_{c+1}\left(\omega_{c+1}^{(k-1)}, \omega_c^{(k-1)}\right) \quad (6)$$

$$\omega_{c+1}^{(k)} = arg\min_{\omega} L_{c+1}(\omega) + \frac{\lambda}{2\alpha_k}\left\|\omega - \mu_{c+1}^{(k)}\right\|^2 \quad (7)$$

EXPERIMENTAL RESULTS AND DISCUSSION

Development environment and datasets

Our deep learning model was constructed using the popular deep learning framework PyTorch, version 1.6.0, along with Python, version 3.7.1. We adopted the identical network configuration as de-scribed in the referenced paper (L. Yu *et al.*, 2022). Specifically, we employed a cyclic federated learning framework, utilizing a Kafka cluster as the medium for exchanging model parameters. The GAN model we used is a conditional Wasserstein Generative Adversarial Network with Gradient Penalty (WGAN-GP). We utilized two distinct medical datasets: the Alzheimer's disease dataset from the Kaggle contest (url: https://www.kaggle.com/datasets/tourist55/alzheimers-dataset-4-class-of-images) and the ADNI MRI dataset (url: https://adni.loni.usc.edu/data-samples/access-data/). The Alzheimer's disease dataset served as the target domain data, while the ADNI dataset was employed as the source domain data for transfer learning in pre-training the GANs. Specifically, The Alzheimer's disease dataset consists of four classes of MRI images in both the training and testing sets, including mild demented, moderate demented, non demented, and very mild demented. We aim to train a general deep learning model via federated learning to be applied in Alzheimer's disease classification tasks. The ADNI MRI image dataset we utilized comprises brain MRI scans from Alzheimer's disease (AD) patients, Mild Cognitive Impairment (MCI) patients, and normal elderly individuals. These images provide detailed information about brain structure, morphology, and pathology. We will employ this dataset to pretrain a WGAN-GP on client-side.

Evaluation

The performance of our algorithm is primarily evaluated based on the classification accuracy. Additionally, we utilize the maximum mean difference (MMD) to quantify the distribution discrepancy be-tween the generated virtual data and the target domain dataset. The squared MMD between two data distributions can be mathematically expressed as:

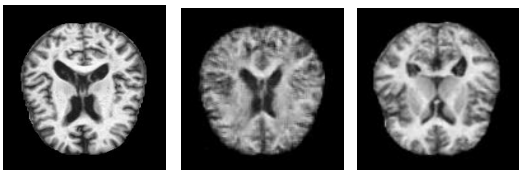
$$MMD^2(x, y) = \|E[\varphi(x)] - E[\varphi(y)]\|^2 \quad (8)$$

Where $\varphi(\cdot)$ denotes the mapping to the regenerated Hilbert space (RKHS).

Results

In CFL_DS_KD, it is essential to ensure that GANs trained on the client's local datasets can adequately learn

the local distribution knowledge and generate high-quality virtual datasets. Due to the small and heterogeneous nature of the local client datasets, it is challenging for the local GANs to fully capture the local distribution knowledge. Therefore, we employ transfer learning to allow GANs to initially learn distribution knowledge from the source domain before fine-tuning them on the local datasets. Fig. 4 demonstrates a comparison of medical images generated using GAN models with and without transfer learning.



(a) (b) (c)

Fig. 4. *(a). An image sample of the target domain dataset. (b). An image sample generated by the GAN generator without transfer learning. (c). An image sample generated by the GAN generator with transfer learning.*

Obviously, we can find that the GAN using transfer learning is better than the original GAN in terms of clarity, contour, texture, etc. of the generated data. Furthermore, we can measure the quality of the data generated by the generator by calculating the MMD value between the generated data and the target domain data.

Table 1. *Distribution discrepancy between generated data and target domain data under different non-iid scenarios.*

Methods	MMD		
	0.5114	0.8630	1.0296
T_GAN-Target Domain	0.5308	0.5858	1.0402
GAN-Target Domain	0.6221	0.7397	1.1279

As shown in the Table 1, the first row of the table presents the client MMD values measured under different non-iid scenarios, while the second row represents the MMD between the data generated by GAN with transfer learning and the target domain data under different non-iid scenarios. The third row shows the MMD between the data generated by GAN without pre-training and the target domain data under different non-iid scenarios. It can be observed that the MMD between the data distribution of the GAN generated through transfer learning and the data distribution of the target domain is smaller. This indicates that the data generated by the GAN with transfer learning is more similar to the target domain data, thus better reflecting the client's data distribution.

The aim of improving the learning performance of client's local distribution is to enable the clients in cyclic federated learning to share their respective real distributions. Therefore, we further evaluate the im-proved strategy from the algorithmic perspective to investigate the impact of transfer learning-based generators on algorithm performance under different client distribution disparities.

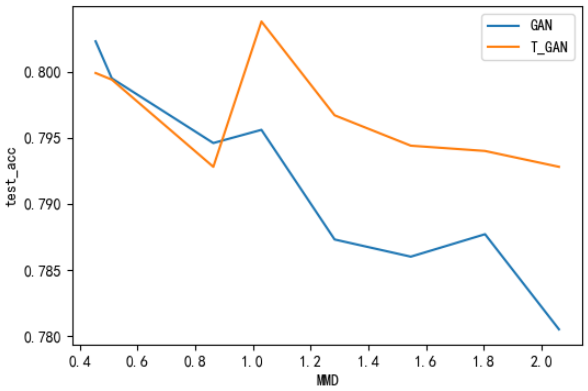


Fig. 5. *The influence of transfer learning GAN and non-transfer learning GAN on performance of CFL_DS_KD was evaluated at different MMD levels.*

Table 2. *The influence of transfer learning-based GAN and non-transfer learning-based GAN on performance of CFL_DS_KD was evaluated at different MMD levels*.

Methods	MMD							
	0.4554	0.5144	0.8630	1.0296	1.2830	1.5468	1.8038	2.0593
GAN	80.23%	79.95%	79.46%	79.56%	78.73%	78.60%	78.77%	78.05%
T_GAN	79.99%	79.84%	79.58%	80.38%	79.67%	79.44%	79.40%	79.28%

As Fig. 5 and Table 2 show, it can be observed that when the MMD is below 0.5, indicating that the client data distributions are very similar, the GANs trained with transfer learning have a negative impact on algorithm performance. When the MMD is be-tween 0.5 and 1.2, indicating that there are some differences in client data distributions but not significant, both methods show similar performance, with a slight advantage for the transfer learning approach. However, when the MMD is greater than 1.2, indicating significant differences in client data distributions, transfer learning shows a noticeable improvement in algorithm performance. Fig. 6 specifically demonstrate the impact of transfer learning-based GAN and non-transfer learning-based GAN on the performance of CFL_DS_KD under different MMD values. It can be observed that the performance of the transfer learning-based GAN is better than the non-transfer learning-based GAN, and this effect becomes more prominent as the MMD increases. When the client distribution disparities are small, the transfer learning-based generator does not provide an advantage. This could be due to the fact that after the GAN learns knowledge from the source domain, fine-tuning on the target domain does not enable the model to adapt well to the target domain distribution, resulting in the model parameters being biased towards the source domain and leading to a deterioration in performance.

Furthermore, to further highlight the advantages of using transfer learning-based GAN, we compared it with other algorithms as shown in the box plot in Fig. 7, where the MMD increases from the top left corner to the bottom right corner. By dynamically increasing the MMD, we can observe that as the MMD increases, indicating more inconsistent data distributions of clients, the performance of the cyclic federated averaging model (CFL_FedAvg) declines rapidly. The non-transfer learning-based GAN per-forms at an intermediate level, while the transfer learning-based GAN exhibits the best and most stable performance. Fig. 8 presents a performance comparison of different methods under different communication rounds. It can also be observed that the transfer learning-based GAN achieves the greatest improvement in the algorithm, and its performance is on par with or even surpasses centralized learning methods.

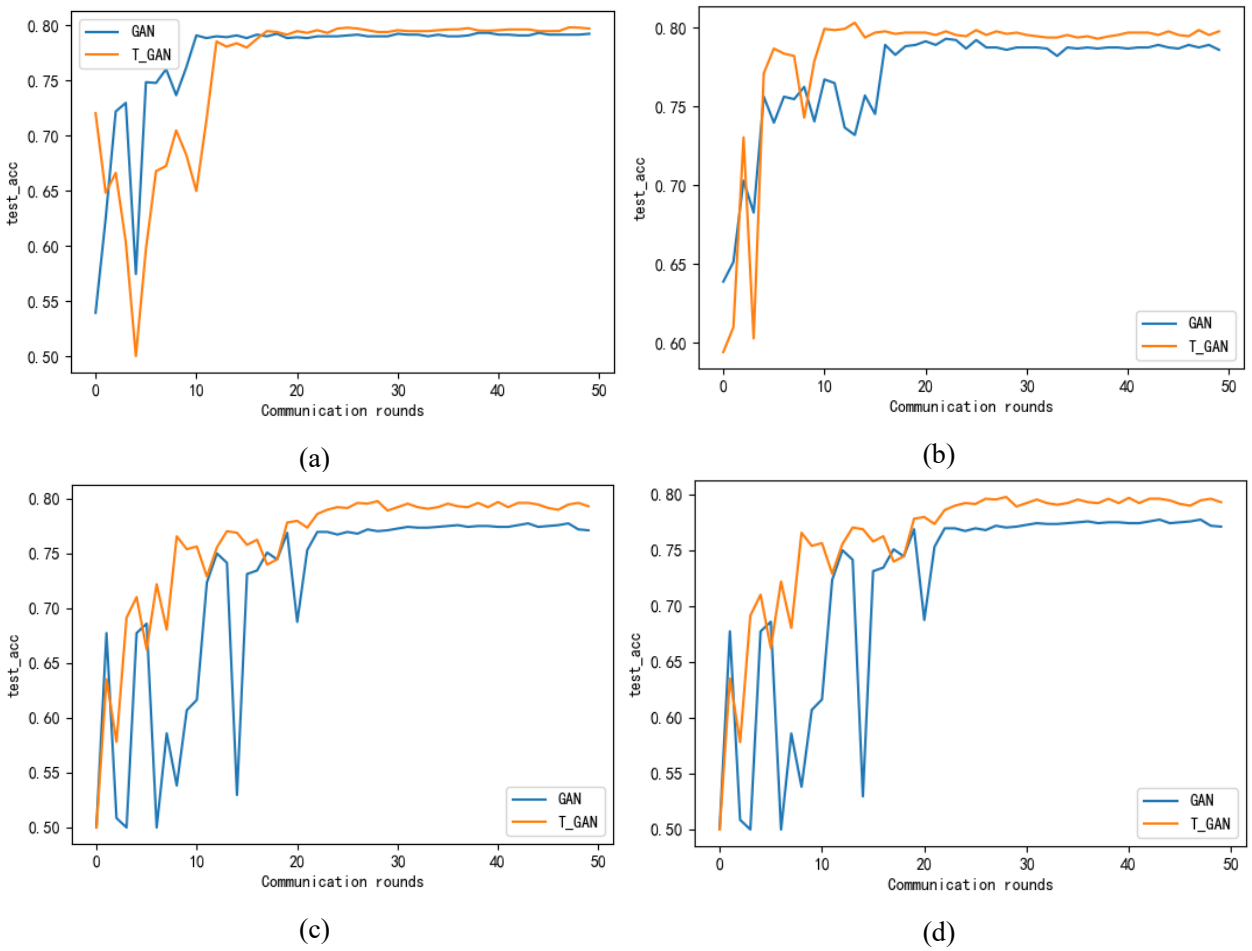


Fig. 6. *The influence of transfer learning GAN and non-transfer learning GAN on performance of CFL_DS_KD was evaluated at different MMD levels. (a) MMD=0.863 (b) MMD=1.546. (c) MMD=1.803. (d) MMD=2.05.*

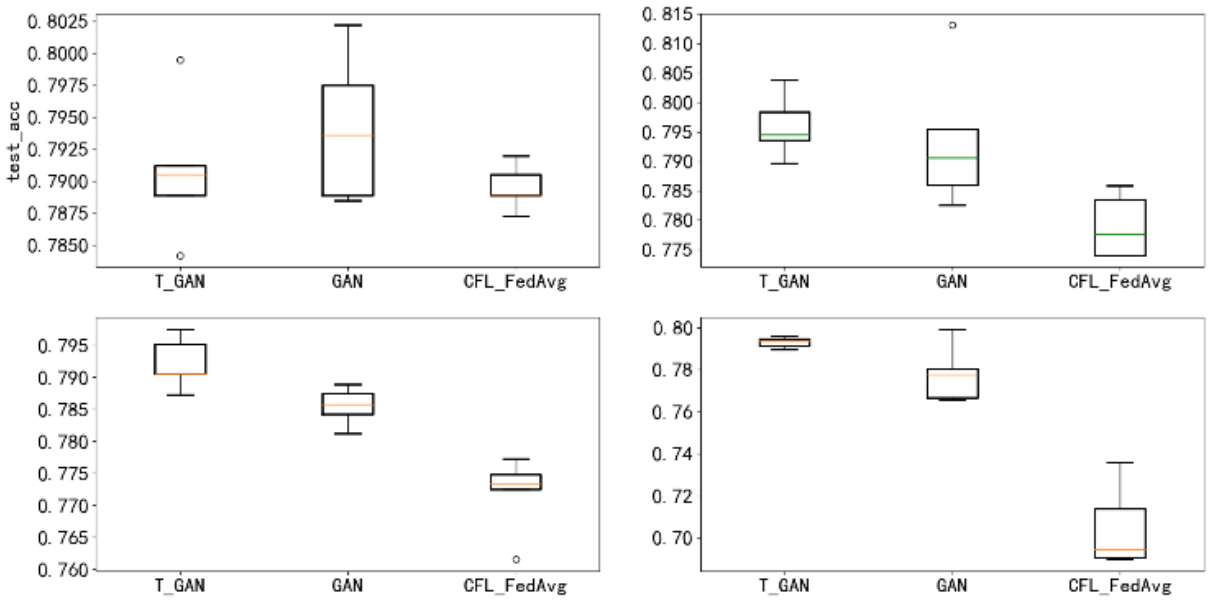


Fig. 7. *Comparison of accuracy box plots for different methods at MMD*

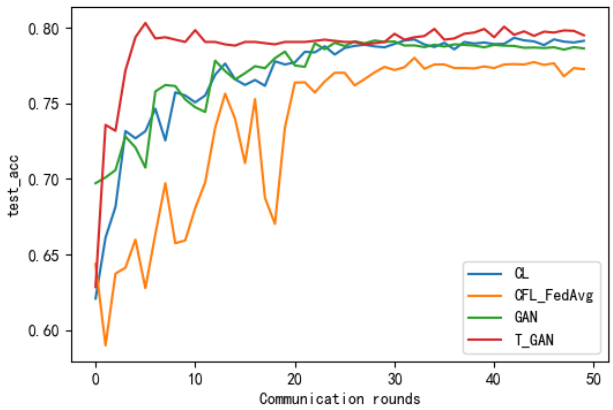


Fig. 8. *Performance comparison of different meth-ods at different communication rounds.*

CONCLUSION

In general, this work focuses on enhancing the learning performance of the client's local data distribution. To overcome the challenges posed by data scarcity and heterogeneous distributions among clients' datasets, we propose the utilization of trans-fer learning to assist GANs in better capturing the underlying data distributions of the clients. Subsequently, the adequately trained GANs are applied within the framework of cyclic federated learning, which incorporates distribution information sharing and knowledge distillation. Through rigorous experimentation and evaluation, we provide evidence of the effectiveness of transfer learning in improving the performance of GANs in learning the client's data distribution, thereby enhancing the overall algorithmic performance.

ACKNOWLEDGMENTS

The work was supported in part by National Natural Science Foundation of China (No. 62171287), Science & Technology Program of Shenzhen (No. JCYJ20220818100004008)

REFERENCES

da Nóbrega RVM, Rebouças Filho PP, Rodrigues MB, da Silva SPP, Dourado Júnior CMJM, de Albuquerque VHC(2020). Lung nodule malignancy classification in chest computed tomography im-ages using transfer learning and convolutional neural networks. Neural Comput Appl 32: 11065-82.

Gessert N, Bengs M, Wittig L, Drömann D, Keck T, Schlaefer A, B.Ellebrecht D(2019). Deep transfer learning methods for colon cancer classification in confocal laser microscopy images. INT J Comput Ass Rad 14: 1837–45.

Hassan M, Ali S, Alquhayz H, Safdar K (2020). Developing intelligent medical image modality classification system using deep transfer learning and LDA. Sci Rep 10: 12868.

Hsu TMH, Qi H, Brown M (2019). Measuring the effects of non-identical data distribution for federated visual classification. arXiv preprint arXiv:1909.06335

Karimireddy SP, Kale S, Mohri M, Reddi S, Stich S, Suresh AT (2020). Scaffold: Stochastic controlled averaging for federated learning. International Conference on Machine Learning: 5132-43.

Li T, Sahu AK, Zaheer M, Sanjabi M, Talwalkar A, Smith V(2020). Federated optimization in heterogeneous networks. MLSys 2: 429-50.

McMahan HB, Moore E, Ramage D, Hampson S, and y. Arcas BA(2017). Communication-Efficient Learning of Deep Networks from Decentralized Data. Artificial Intelligence and Statistics: 1273-82.

Miotto R, Wang F, Wang S, Jiang X, and Dudley JT (2017). Deep learning for healthcare: review, opportunities and challenges. Brief Bioinform 19: 1236-46.

Pan S, Yang Q (2010). A Survey on Transfer Learning. IEEE T on Knowl Data En 22: 1345-59.

Shoham A, Avidor T, Keren A, Israel N, Benditkis D, Mor-Yosef L, Zeitak I (2019). Overcoming forget-ting in federated learning on non-iid data. 31st Conference on Neural Information Processing Systems .

Swati ZNK, Zhao Q, Kabir M, Ali, F, Ali Z, Ahmed S, Lu J(2019). Brain tumor classification for MR images using transfer learning and fine-tuning. Comput Med Imag Grap 75: 34-46.

Wang H, Yurochkin M, Sun Y, Papailiopoulos D, Khazaeni Y (2020). Federated learning with matched averaging. International Conference on Learning Representations 2020.

Wang J, Liu Q, Liang H, Joshi G, Poor HV (2020). Tackling the objective inconsistency problem in heterogeneous federated optimization. 34th Conference on Neural Information Processing Systems 33: 7611-23.

Xiao J, Du C, Duan Z, Guo W (2021). A novel serv-er-side aggregation strategy for federated learning in non-iid situations. 20th International Symposium on Parallel and Distributed Computing (ISPDC): 17-24.

Yang Q, Liu Y, Chen T, and Tong Y (2019). Federated Machine Learning: Concept and Applications. ACM T Intel Syst Tec 10: 1-19.

Yu L and Huang J (2022). Cyclic Federated Learning Method Based on Distribution Information Sharing and Knowledge Distillation for Medical Data. Electronics 11: 4039.