B^2-FedGAN: Balanced Bi-directional Federated GAN

Ali Anaissi[1][0000−0002−8864−0314] and Basem Suleiman[1,2][0000−0003−2674−0253]

[1] School of Computer Science, University of Sydney, Australia
[2] School of Computer Science and Engineering, The University of New South Wales, Sydney, Australia
basem.suleiman@unsw.edu.au
{ali.anaissi, basem.suleiman}@sydney.edu.au

Abstract. In Federated Learning (FL), a shared model is learned across dispersive clients each of which often has small and heterogeneous data. As such, datasets in FL setting may suffer from the non-IID (Independent and identically distributed) problem. In this paper, we propose a BAGAN as machine learning model which has the ability to create data for minority classes, and a Bi-FedAvg model as a new approach to mitigate non-IID problems in FL settings. The performance comparison between FedAvg and Bi-FedAvg in both IID and Non-IID environments will be shown in terms of accuracy, converge stability and category cross-entropy loss. On the other hand, the training and testing performance among FedAvg, FedAvg with a conditional GAN model, and FedAvg with BAGAN-GP model, on IID and Non-IID environments with three imbalanced datasets will be compared and discussed. The results indicate that Bi-FedAvg fails to outperform Fed-Avg, for Bi-FedAvg suffers from model quality loss or even divergence when running on non-IID data partitions. In addition to that, our experiments demonstrate that higher quality images for complex image datasets can be generated by BAGAN and combining federated learning and Balancing GAN model together is conducive to obtaining a high-level privacy-preserving capability and achieving more competitive model performance. The project will give an inspired further exploration of the implementation of a combination between Federated learning and BAGAN on image classification in real-world scenarios.

Keywords: GAN; Federated Learning; Imbalanced dataset

1 Introduction

Deep learning methods have been employed extensively in the area of image classification. Due to its complexities, deep learning models require a large number of samples for learning process. In such domain such as medical applications it is often not possible to collect large datasets in one single hospitals (a.k.a clients). At the same time, it's also hard to combine data from different clients at one

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-35995-8_27

single central location because of their privacy regulations [7]. In this sense, Federated Learning (FL) became a dominant solution to this kind of problem [9]. In Federated Learning, only local client data will be utilized to train a local model rather than be gathered to a central server, and the central server will subsequently aggregate only the local model coefficients into the global model. The privacy of users is well safeguarded in Federated Learning since no other parties will ever have access to client data.

Nevertheless, in real situations, data from different clients are very heterogeneous. In other words, data of different classes are not evenly distributed in multiple clients [3]. The clients may have too few samples of certain classes, or missing some classes. This is the non independent and identically distributed, or Non-IID data problem which is currently an open research question in federated learning. For instance, the predicting accuracy of standard federated learning models is lower than 55% with highly skewed Non-IID data, where each client only contains one class of data, compared with IID type of data [20].

This paper proposes a novel federated learning method based on a balance GAN (BGAN) model to handle the problem of imbalanced classes in datasets, and a bi-directional FedAvg method to mitigate the effect of Non-IID problem. The adversarial module in BGAN will learn the pattern of both majority and minority classes in the dataset, and its generative model will generate images for the minority classes.

The rest of this paper is organized as follows. In Section 2, various related studies are discussed. In Section 3, we present our approach and data preprocessing methodology. Section 4 presents our experiment and results. Finally, the conclusions are drawn in Section 5.

2 Related Work

Many studies underscore the potential of deep learning will help to identify complex patterns in medical industry. For that, sufficiently large and diverse datasets are required for training. However, as multi-institutional collaborations centrally share the patient data and as a result, there will be privacy and ownership challenges [17]. Federated Learning have been recently considered as a potential solution for building a predictive model in medical industry without sharing patient data to the Central model [14].

FedAvg algorithm is the original version of federated aggregation algorithm which was initially proposed by McMahan et al.[19]. It uses the local SGD updates to build a global model by taking average model coeffcients from a subset of clients with non-IID data. In FedAvg, a central server is used to communicate between clients without accessing the data in local server. For Federated Learning, three main challenges have attracted several researches to improve the FedAvg algorithm. (1) how to handle non-identically distributed (Non-IID) data across the network (statistical heterogeneity); (2) how to aggregate the coeffcient at the central model; and (3) who and when to communicate client's model coefficients. Accordingly, several derived models of FedAvg have been developed

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-35995-8_27

with theoretical guarantee. For instance, a method called FedProx is proposed by LiTian [18] to tackle the Non-IID problem. FedProx can be viewed as a generalization and re-parametrization of FedAvg. Despite very little changes to the method itself, but there will be significant implications both in theory and in practice. Compared with FedAvg, one key improvement for FedProx is that it introduces an additional proximal term to the local training, which essentially restricts the local updates to be closer to the latest global model, and that will help the federated training to converge faster [18]. Subsequently, pFedBayes [19] is also proposed to mitigate the non-Non-IID problem. The idea for this model is that each client uses the aggregated global distribution as prior distribution and updates its personal distribution by balancing the construction error over its personal data and the KL divergence with aggregated global distribution. Another algorithm called PC-FedAvg [2, 1], it personalizes the resulting support vectors to addresses the problem of Non-IID distribution of data in FL. Simlarly, pFedMe [6] formulates a new bi-level optimization problem by using the Moreau envelope as a regularized loss function. The idea behind that is allowing clients to pursue their own models with different directions, but not stay far away from the reference point. In this sense, the statistical diversity among clients will be smaller. Therefore, the above algorithms could help us to think about how to improve the current model or make some changes into the model for the Non-IID problem and ultimately get a better performance.

A very recent method propsed by LiZheng [12] called FedFocus. The idea behind this method is that the training loss of each model is taken as the basis for parameter aggregation weights, and as training layer deepens, a constantly updated dynamic factor is designed to stabilize the aggregation process, which could improve the training efficiency and accuracy. On the other hand, the paper applies it into COVID-19 detection on CXR images. The author claimed that the training efficiency, accuracy and stability was significantly improved by using FedFocus. In this paper the authors utilise a GAN model on client side for learning process. However, datasets are often class-imbalanced, which will negatively affect the accuracy of deep learning classifiers [13]. It is pointed that the existing GANs and their training regimes only work well on balanced datasets but fail to be effective in case of imbalanced datasets.

This paper proposes a bi-directional version of the traditional FedAvg model to alleviate the problem of Non-IID data with a better aggregation method on the client side. Furthermore, our method used an improved version of GAN which has the ability to deal with class-imbalanced datasets.

3 Bi-directional FedAvg Based Balanced GAN for Learning Process

Federated learning uses a distributed framework which allows multiple clients to collaboratively train a machine learning model using their own unique dataset. This is to say that the client trains local data to obtain a local model and aggregates the local model by updating parameters to the central server to obtain

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-35995-8_27

a global model. After multiple iterations until meeting a terminating condition, the final model converges to the centralised machine learning result. Based on the research work we reviewed, the accuracy of model decreases when the diversity of data acroos the clients increases [5]. Hence, our new method i.e Bi-FedAvg (Bi-directional FedAvg) adds an extra computation layer on the client side after the central server has published the aggregated global model. Figure 1 presents our Bi-FedAvg framework architecture.

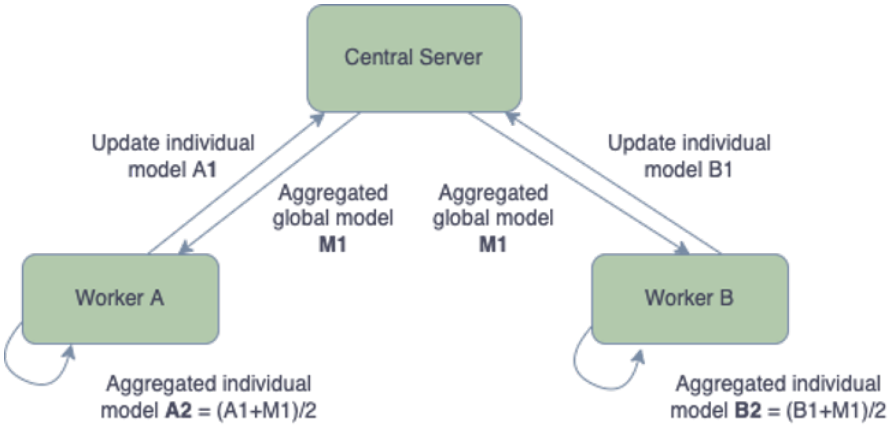

Fig. 1: Core Architecture of Bi-FedAvg.

As can be seen from Figure 1, Bi-FedAvg averages the current aggregated weights from the global model with the local client model's weights to alleviate the problem that training data differs from user to user as well as not population-specific characteristics.

We implement a sample MLP model which includes two fully-connected layers and one softmax layer. After the training iterations, it is expected that local data contributions will become less divergent. However, it still cannot mitigate the Non-IID problem, which means the model may be negatively affected by Non-IID data among clients. In reality, the majority of clients' data categories are unevenly distributed. Thus, we will use the BAGAN to solve the imbalanced dataset. The BAGAN is a methodology to restore the balance of an imbalanced dataset by using generative adversarial networks. Additionally, it has a higher accuracy of deep-learning classifiers trained over the augmented dataset where the balance has been restored [15]. However, it might also suffer from some problems. For instance, it is unstable when images in different classes look similar. For example, the imbalanced Flowers dataset has many similar classes so BAGAN performs not well. On the other hand, it is hard to train and sensitive to its architecture and hyper-parameters [8]. Therefore, we have adopted the improved version of BAGAN-GP, which can improve the loss function of BAGAN with gradient penalty. Moreover, Huang & Jafari (2021) propose an

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-35995-8_27

architecture of autoencoder with an intermediate embedding model, which helps the autoencoder learn the label information directly. In our model architecture, the BAGAN model will run on each client to augment the local data and restore class balance, then federated learning will be applied to the combination of augmented data and original data of the client.

4 Experimental Setup

4.1 Data Collection

We conduct three different experiments using three different datasets i.e MNIST, CIFAR-10 and COVID-19 Radiography. The MNIST dataset was published by Lecun team [11], it consists of a picture for a handwritten number and a corresponding label. There are 10 categories of images, corresponding to 10 numbers from 0 to 9. The MNIST dataset contains 60000 training sets and 10000 testing sets. As for the CIFAR-10, it is a labelled subset of 80 million tiny images dataset which was collected by Krizhevsky, Nair and Hinton [10]. The CIFAR-10 dataset contains 60000 images with a 50000 training set and a 10000 testing set, which include 10 labelled classes: airplane, automobile, bird, cat, deer, dog, frog, horse, ship and truck. The COVID-19 Radiography dataset is a chest X-ray images about COVID-19 positive cases along with normal and viral pneumonia images created by a team of researchers along with medical doctors [4] [16]. This dataset contains 3616 COVID-19 images, 10192 normal, 6012 lung opacity and 1345 viral pneumonia images. Figure 2 shows the distribution of classes in COVID-19 dataset. As can be seen, the COVID-19 dataset is highly imbalanced. The Normal class has 10192 images, which is 3 times as many as COVID-19 class (3616). The details of three datasets, including image resolution, number of classes, and minimum/maximum/total number of images in classes, were shown in Table 1

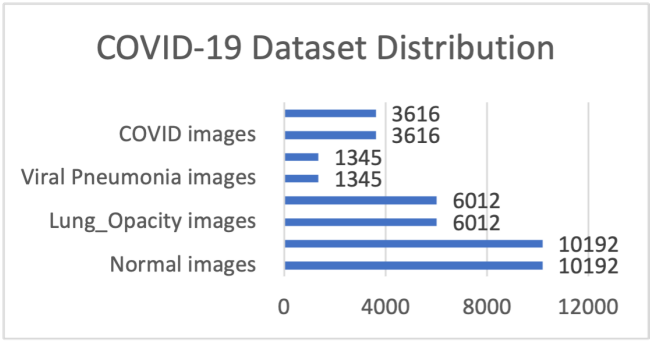

Fig. 2: The distribution of classes in COVID-19 dataset .

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-35995-8_27

Table 1: Datasets information including resolution, number of classes, and min/max/total number of images in classes.

Dataset	Resolution	Classes	Min	Max	Total
MNIST	28 * 28	10	5421	6742	60000
CIFAR-10	32 * 32	10	5000	5000	50000
COVID-19	299 * 299	4	1345	10192	21165

4.2 Results

Comparison experiments on FedAvg and Bi-FedAvg: The first experiment we conducted was to evaluate the performance of Bi-FedAvg algorithm in both IID and Non-IID settings, and to compare the resultant accuracies with FedAvg using the three different image datasets MNIST [11], CIFAR-10 [10] and COVID-19 Radiography dataset. Initially, we apply the following data pre-processing procedures on the three datasets. RGB images in CIFAR-10 and COVID-19 datasets were converted to grayscale. Images in the COVID-19 dataset were resized to 64 * 64 pixels to avoid large time consumption and insufficient memory problems in the testing environment. After scaling feature vectors to 0 to 1 scale, training and test datasets with a test size of 0.30 were created. Using RELU as the activation function, a multilayer perceptron with two hidden layers and 200 hidden units on each layer was constructed. Each client contained the same number of data points. However, clients in the IID and Non-IID settings had different arrangements of the data. The data was further processed and batched with a batch size of 32 for each client. For FedAvg, an initial global model's weight was set and would serve as the initial weights for all local models. In each communication round for a total of 30 to 300 communication rounds, 10 randomly selected clients were fitted into local models. Local weights for each client were averaged, updated to a global model, and this process was iterated to find the performance on the test dataset. For Bi-FedAvg, an additional average process was implemented between the global model weight and local models' weights before assigning the global weight to each local model.

The experiment was implemented in both IID and Non-IID settings. IID-imbalanced setting indicates that each client has an imbalanced dataset containing all classes. Non-IID setting means that each client has an imbalanced dataset with a random number of classes obtained. Figure 3 showed the training accuracy and categorical cross entropy loss plot on three datasets in the IID environment. It was evident that Bi-FedAvg had upward trends for training accuracy performance on all datasets, however FedAvg only had an upward trend for training accuracy performance on the MNIST dataset, whereas CIFAR-10 and COVID-19 datasets exhibited a brief upward trend for 10 communication rounds followed by a decline trend. The early stopping callback approach was used to terminate training when the validation loss reached a stopping decrease;

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-35995-8_27

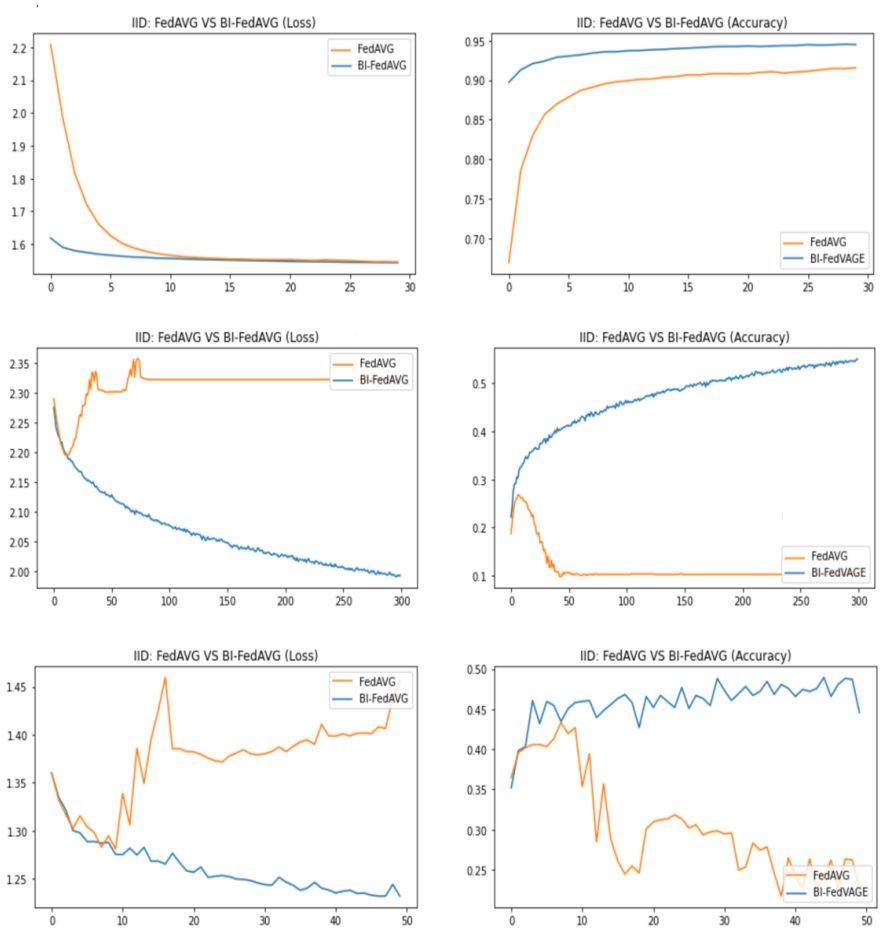

Fig. 3: Training accuracy and categorical cross entropy loss plot of three datasets in IID environment. Top: MNIST. Middle: CIFAR-10. Bottom: COVID-19 dataset.

nevertheless, the performance of Bi-FedAvg outperformed FedAvg on all selected datasets in the IID setting.

Experiments on Bi-FedAvg with CGAN, BAGAN-GP, and without GAN model: The second experiment was designed to assess the training and testing performance of several experiment settings, including FedAvg, FedAvg with a conditional GAN model, and FedAvg with BAGAN-GP model, on IID and Non-IID environments with three imbalanced datasets. This experiment utilized the same datasets as the previous one, including the MNIST, CIFAR-

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-35995-8_27

10, and COVID-19 Radiography datasets. We applied the same pre-processing and the structure of multilayer perceptrons as in the first experiment. Before distributing data samples into each client, BAGAN-GP or CGAN were used to restore class balance by generating images for classes with insufficient numbers of images. The data were then combined with real and generated images to create additional data for each client, which was subsequently fitted into local models. The hyperparameters for BAGAN-GP are defined as follows: The optimizer was set to the Adam algorithm with a learning rate of 0.0002 and momentum (0.5, 0.9); the default batch size and dimension of the latent vector were set to 128, and the training ratio of the discriminator to the generator was set to 5 [8]. Each client's images were fitted to BAGAN-GP with 100 learning steps and two epochs per learning step. For CGAN, the discriminator and generator were set to multilayer perceptron with three hidden layers and LeakReLU as activation function. The optimizer was the same as BAGAN-GP and the batch size was set to 512. To restore balance, a random half-batch of images was selected for each epoch, and 2000 epochs were run to generate images. For the IID environment, an additional parameter was set to the randomly selected step to allocate each client with data points from all classes with a proportion of minimum data points in each class. This ensured that there were no situations where certain classes only contain one or two data points.

Figure 4 shows the training accuracy and category cross-entropy loss of several experiments using three datasets in an IID environment. For the performance on the MNIST and COVID-19 datasets, all three experiments had similar training and testing performances. The accuracy plots on the right-hand side increased rapidly on the first 50 communication rounds and turned to increase mildly after the rest of the communication rounds. FedAvg and FedAvg with BAGAN-GP performed slightly better on test datasets, achieving 96.967% accuracy and 1.496 category cross-entropy loss on the MNIST dataset, while FedAvg with CGAN performed slightly better with 81.17% accuracy and 0.931 category cross-entropy loss on COVID-19 dataset. On the Cifar-10 dataset, the growth trend of accuracy plots was relatively more gradual than that on the MNIST dataset. FedAvg with BAGAN-GP had the best performance on the test dataset with 42.41% accuracy and 2.052 categorical cross-entropy loss. Table 2 demonstrated the test accuracy and category cross-entropy loss of three experiments using three datasets in the IID environment.

Table 2: Test accuracy and category cross-entropy loss of three experiments using three datasets in the IID environment.

	MNIST-Accuracy	**MNIST-Loss**	**CIFAR-10-Accuracy**	**CIFAR-10-Loss**	**COVID-Accuracy**	**COVID-Loss**
FedAvg	96.97%	1.496	41.15%	2.062	80.42%	0.936
FedAvg+CGAN	96.83%	1.498	40.15%	2.070	81.17%	0.931
FedAvg+BAGAN-GP	96.97%	1.496	42.41%	2.052	80.79%	0.937

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-35995-8_27

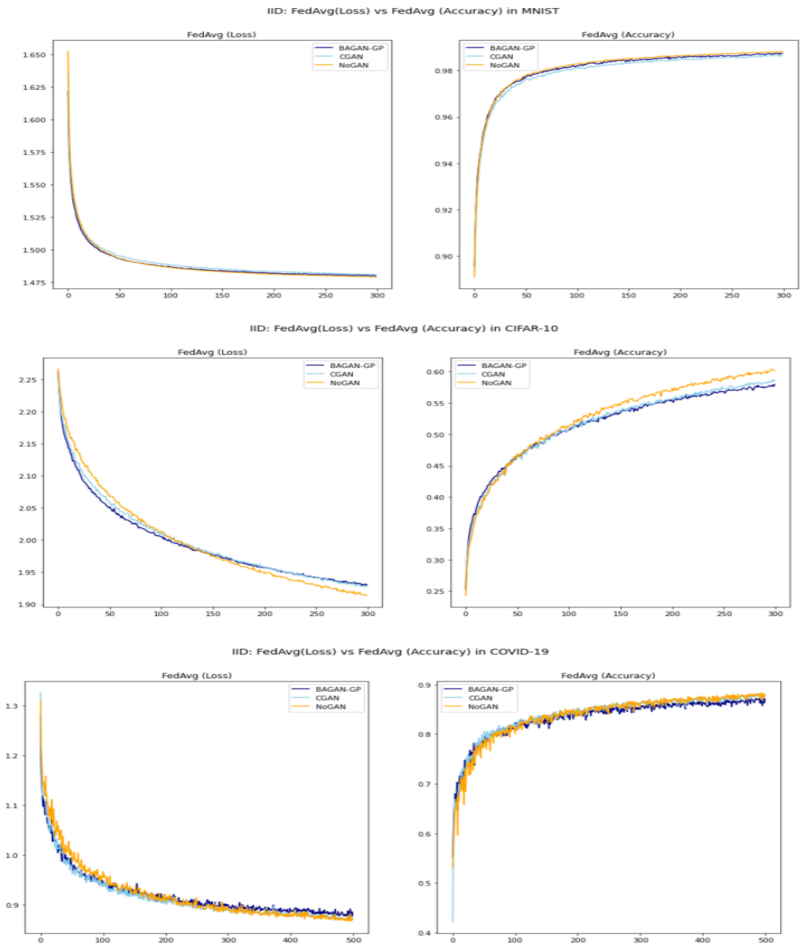

Fig. 4: Training accuracy and category cross-entropy loss of three experiments using three datasets in the IID environment.

For Non-IID environments, a random selection function was assigned to each client in order to distribute data points from a random number of classes with a minimum percentage of data points in each class.

Figure 5 depicted the training accuracy and category cross-entropy loss of several experiments using three datasets in the non-IID environment. For the performance on the MNIST dataset, the training and testing performances of all three experiments are similar. Before 50 communication rounds, the curve displayed a few mild undulations. FedAvg with the BAGAN-GP model gets the greatest performance on the test dataset, with 96.23 percent accuracy and 1.508 percent category cross-entropy loss. On the CIFAR-10 dataset, the performance plots of all training experiments demonstrated large fluctuations, but

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-35995-8_27

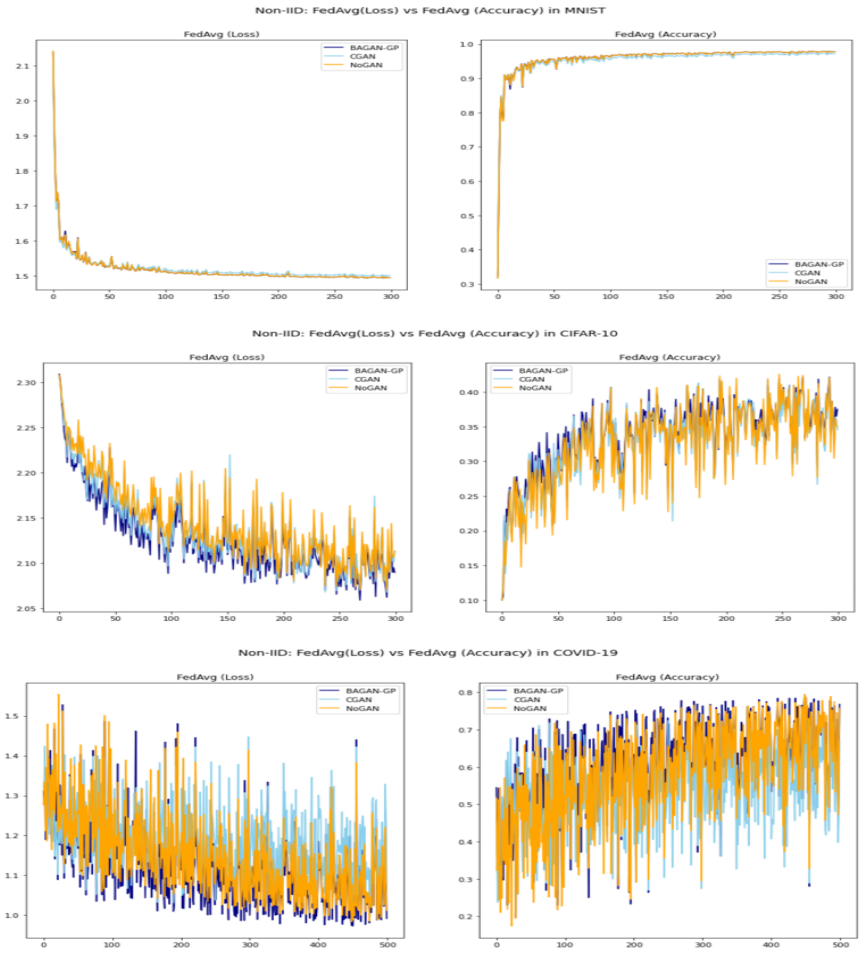

Fig. 5: Training accuracy and category cross-entropy loss of three experiments using three datasets in the non-IID environment

an overall upward trend. FedAvg with the BAGAN-GP model achieved the best performance on the test dataset, with 32.29% accuracy and 2.131 category cross-entropy loss. On the COVID-19 dataset, the performance plots of all training experiments demonstrated an extreme fluctuation, which indicated that the model did not converge within 500 communication rounds. On the test dataset, FedAvg performed the best, followed by FedAvg with BAGAN-GP and FedAvg with CGAN. FedAvg and FedAvg with BAGAN-GP had a similar upward trend; however, FedAvg with BAGAN-GP had a greater peak for the same variation during the training process as FedAvg without GAN model, indicating that BAGAN-GP should have a superior performance. Possible mistakes or volatility in the model's learning processes may have impacted the final accuracy or

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-35995-8_27

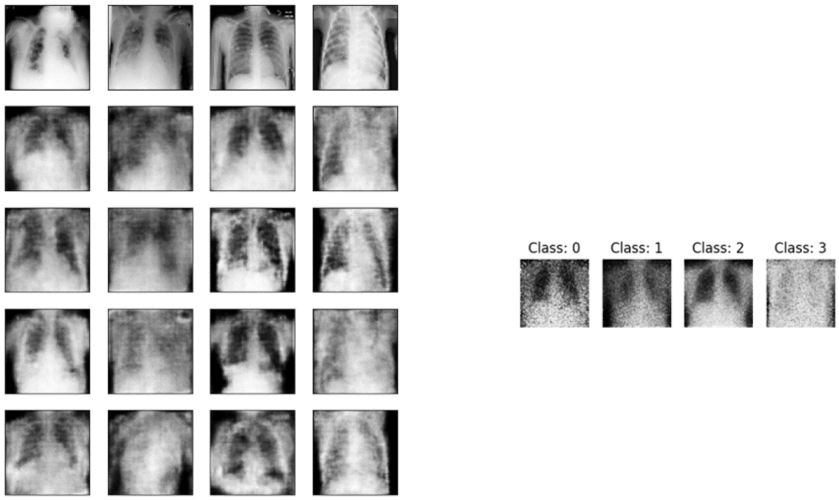

Fig. 6: Comparison of generated images for BAGAN-GP and CGAN model in COVID-19 dataset. Left: Generated images for BAGAN-GP. Right: Generated images for CGAN.

category cross-entropy loss on test datasets, which led to BAGAN-GP's poorer performance. Table 3 demonstrated the test accuracy and category cross-entropy loss of three experiments using three datasets in the non-IID environment.

Table 3: Test accuracy and category cross-entropy loss of three experiments using three datasets in the non-IID environment.

	MNIST-Accuracy	**MNIST-Loss**	**CIFAR-10-Accuracy**	**CIFAR-10-Loss**	**COVID-Accuracy**	**COVID-Loss**
FedAvg	96.17%	1.508	29.04%	2.156	73.30%	1.028
FedAvg+CGAN	95.64%	1.513	29.05%	2.158	65.53%	1.090
FedAvg+BAGAN-GP	96.23%	1.508	32.29%	2.131	67.47%	1.063

Furthermore, Figure 6 demonstrated the comparison of generated images for the BAGAN-GP and CGAN model in the COVID-19 dataset. It was evident that BAGAN-GP generated higher-quality images than CGAAN, especially when images had minor differences between classes. The conditional GAN model did not help to increase the performance due to its relatively simple architecture that could not generate high quality images for complex image datasets.

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-35995-8_27

5 Conclusion

We have also introduced BAGAN as a data augmentation tool to resolve the heterogeneous and imbalanced data distribution problem in federated learning. The purpose of using BAGAN is to restore class balance in client data and improve local data quality. Experiment results showed that BAGAN outperforms our baseline model and CGAN in most of the scenarios. Different from other data augmentation methods in federated learning where client data are shared with the server to train a good generator, our GAN model is trained specifically on each client using only client data. While the majority of data augmentation methods used data sharing and may raise the risk of data privacy leakage, our approach can protect client data privacy because no local data or labelling information is shared with the central server. Our research offers a new data augmentation framework for the generative adversarial network (GAN) based method that can enhance model performance while maintaining client data privacy. In the field of bio-medics, gathering and distributing large amounts of medical photographs appears to be impossible, in part because of the lack of sufficient public access to sensitive data and patient privacy concerns. However, research has found that most patients are willing to volunteer their data for research purposes if sufficient precautions have been taken to protect their privacy. Therefore, our framework provides a potential solution to break the boundary of data sharing limitations without leaking patient-sensitive data.

Acknowledgement

We would like to thank Zhibin Ye, Zekai Zhang, Yinxuan Ding, and Yiting Li who helped in carrying out data analysis and model implementation in this project.

References

1. Anaissi, A., Suleiman, B., Alyassine, W.: Personalised federated learning framework for damage detection in structural health monitoring. Journal of Civil Structural Health Monitoring pp. 1–14 (2022)
2. Anaissi, A., Suleiman, B., Alyassine, W.: A personalized federated learning algorithm for one-class support vector machine: An application in anomaly detection. In: Computational Science–ICCS 2022: 22nd International Conference, London, UK, June 21–23, 2022, Proceedings, Part IV. pp. 373–379. Springer (2022)
3. Cao, L.: Beyond i.i.d.: Non-iid thinking, informatics, and learning. IEEE Intelligent Systems **37**(4), 5–17 (2022). https://doi.org/10.1109/MIS.2022.3194618
4. Chowdhury, M., Rahman, T., Khandakar, A., Mazhar, R., Kadir, M., Mahbub, Z., Islam, K., Khan, M.S., Iqbal, A., Al-Emadi, N., Reaz, M.B.I., Islam, M.: Can ai help in screening viral and covid-19 pneumonia? IEEE Access **8**, 132665 –132676 (07 2020). https://doi.org/10.1109/ACCESS.2020.3010287

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-35995-8_27

5. Deng, Y., Kamani, M.M., Mahdavi, M.: Adaptive personalized federated learning (2020). https://doi.org/10.48550/ARXIV.2003.13461, https://arxiv.org/abs/2003.13461
6. Dinh, C.T., Tran, N.H., Nguyen, T.D.: Personalized federated learning with moreau envelopes (2020). https://doi.org/10.48550/ARXIV.2006.08848, https://arxiv.org/abs/2006.08848
7. Ha, T., Dang, T.K., Le, H., Truong, T.A.: Security and privacy issues in deep learning: A brief review. SN Comput. Sci. **1**, 253 (2020). https://doi.org/10.1007/s42979-020-00254-4
8. Huang, G., Jafari, A.: Enhanced balancing gan: minority-class image generation. Neural Computing and Applications (06 2021). https://doi.org/10.1007/s00521-021-06163-8
9. Kaissis, G., Makowski, M.R., Rückert, D., Braren, R.F.: Secure, privacy-preserving and federated machine learning in medical imaging. Nature Machine Intelligence **2**, 305–311 (2020). https://doi.org/10.1038/s42256-020-0186-1
10. Krizhevsky, A., Hinton, G.: Learning multiple layers of features from tiny images. Tech. Rep. 0, University of Toronto, Toronto, Ontario (2009)
11. Lecun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proceedings of the IEEE **86**(11), 2278–2324 (1998). https://doi.org/10.1109/5.726791
12. Li, Z., Xu, X., Cao, X., Liu, W., Zhang, Y., Chen, D., Dai, H.: Integrated cnn and federated learning for covid-19 detection on chest x-ray images. IEEE/ACM Transactions on Computational Biology and Bioinformatics pp. 1–11 (2022). https://doi.org/10.1109/TCBB.2022.3184319
13. Li, Z., Shao, J., Mao, Y., Wang, J.H., Zhang, J.: Federated learning with gan-based data synthesis for non-iid clients (2022). https://doi.org/10.48550/ARXIV.2206.05507, https://arxiv.org/abs/2206.05507
14. Liu, Y., Zhang, L., Ge, N., Li, G.: A systematic literature review on federated learning: From a model quality perspective (2020). https://doi.org/10.48550/ARXIV.2012.01973, https://arxiv.org/abs/2012.01973
15. Mariani, G., Scheidegger, F., Istrate, R., Bekas, C., Malossi, C.: Bagan: Data augmentation with balancing gan (2018). https://doi.org/10.48550/ARXIV.1803.09655, https://arxiv.org/abs/1803.09655
16. Rahman, T., Khandakar, A., Qiblawey, Y., Tahir, A., Kiranyaz, S., Kashem, S.B.A., Islam, M.T., Maadeed, S.A., Zughaier, S.M., Khan, M.S., Chowdhury, M.E.H.: Exploring the effect of image enhancement techniques on covid-19 detection using chest x-rays images (2020). https://doi.org/10.48550/ARXIV.2012.02238, https://arxiv.org/abs/2012.02238
17. Sheller, M.J., Edwards, B., Reina, G.A., Martin, J., Pati, S., Kotrotsou, A., Milchenko, M., Xu, W., Marcus, D., Colen, R.R., Bakas, S.: Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data. Scientific Reports **10** (2020). https://doi.org/10.1038/s41598-020-69250-1
18. Su, L., Xu, J., Yang, P.: A non-parametric view of fedavg and fedprox: Beyond stationary points (2021). https://doi.org/10.48550/ARXIV.2106.15216, https://arxiv.org/abs/2106.15216

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-35995-8_27

19. Zhang, X., Li, Y., Li, W., Guo, K., Shao, Y.: Personalized federated learning via variational bayesian inference. In: International Conference on Machine Learning. pp. 26293–26310. PMLR (2022)
20. Zhao, Y., Li, M., Lai, L., Suda, N., Civin, D., Chandra, V.: Federated learning with non-iid data (2018). https://doi.org/10.48550/ARXIV.1806.00582, https://arxiv.org/abs/1806.00582

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-35995-8_27