JOURNAL OF
EMERGING INVESTIGATORS

Article

Collaboration beats heterogeneity: Improving federated learning-based waste classification

Jiwon Chong[1], Jaehyok Chong[2]
[1]Seoul Scholars International Math and Science, Seoul, South Korea
[2]SAP Labs Korea, Seoul, South Korea

SUMMARY

Classifying waste is a simple but useful exercise to protect the environment. Based on the success of deep learning, recent works have attempted to develop a waste classification model using deep neural networks. As we attempted to make our very own model to address waste classification, we recognized a common issue that other researchers reported as well: the lack of training data. This is an especially big problem when attempting to train a model for a task that has not been commonly explored before. Further research led us to federated learning (FL) for a solution, as it allows participants to aid in training the model using their own data. However, instead of sending the data from the clients to the server for training, FL trains the model on the clients' machines using the clients' data, and then aggregates the trained models into the server's model to protect the clients' private information. This means that there is a higher chance that the data used to train local models are heterogeneous, decreasing the overall performance. To overcome this issue, we ran a multitude of tests to determine how to maximize performance. For these tests, we hypothesized that the impact of data heterogeneity on an FL framework can be diminished through increasing the client participation ratio. Then, we measured the accuracies of the models trained with varying data heterogeneity, participation ratio, and the number of clients. From the results, we discovered that with less clients, having a higher participation ratio resulted in less accuracy degradation by the data heterogeneity.

INTRODUCTION

Two-hundred and ninety-two million tons represents the amount of waste produced by the United States in 2018, and only 32.1% of this waste was recycled or composted (1). This recycled waste shows a slight decrease from the 35% measured in 2017 due to the recycling capabilities being unable to keep up with the increase in waste (1). As the amount of waste increases, recycling capacity needs to increase along with it to help overcome the increasing waste production. However, current methods require human labor, and therefore are not scalable (2, 3). Thus, we decided to design a neural network that could help classify waste and make recycling much easier.

With deep learning becoming the new trend, many researchers are designing and training their own neural network models, with various objectives. One of the applications of deep learning that is being explored is waste classification (4-6). There is one problem, however, since training a model from scratch requires a tremendous amount of data. This is an especially big issue when attempting to tackle a task that has not been explored before, as there is an especially small amount of public data available. To counteract this data predicament, various researchers have introduced transfer learning with pre-trained models (4-6). Researchers first used a pre-trained residual network (ResNet) model with 50 layers (ResNet-50) to build an intelligent waste classification system (4). Another group of researchers used a pre-trained ResNeXt model to improve the predictive performance (5). Others proposed federated learning (FL) for training waste classification models using the pre-trained ResNet model (6). They also applied active learning to overcome the necessity of qualified annotation on the clients' training data (6).

A convolutional neural network (CNN) is a kind of neural network which is most commonly used to analyze visual images. As the number of layers increases in CNN, the parameters of the deeper layers start approaching zero. To handle this problem, ResNet has 'shortcut' connections around every two layers, forming a 'residual block.' The resulting structure has the effect of reducing the depth of the network by skipping over the layers within the blocks (7). ResNeXt—aggregated residual transformations network—inherits the ResNet and adopts the split-transform-merge (aggregate) strategy of the inception model in each residual block. For splitting and aggregation, the model exposes a new dimension, which is called cardinality (the number of branches to be split) or the size of the set transformations (8). These two models are popular in image classification tasks because the pre-trained versions of the models are publicly available, and researchers can utilize the models for their own objectives by fine-tuning them.

Transfer learning is an efficient method to get better performance even when the training data is not enough by using a pre-trained model as the starting checkpoint for a new task. For the image classification task, it is common to use the checkpoint trained with a large-scale dataset, such as ImageNet (9). Normally, pre-trained models are used in the same domain tasks, but it is also possible to transfer the knowledge between different domain tasks. To utilize this effect of transfer learning, we used a ResNet with 18 layers (ResNet-18) as our pre-trained model and further fine-tuned using the FL scheme with the waste classification dataset.

FL was first proposed by McMahan *et al*. as an algorithm for training the global model with the client (or edge device) data without leaking any private and potentially compromising data (10). These authors summarize the three constraints of FL as data security, data heterogeneity, and limitation of communication bandwidth, which impose restrictions on FL capabilities. Under these constraints, a federated averaging (FedAvg) algorithm was proposed, which trains the global model by averaging the model parameters locally trained in each client and applying the averaged parameters to the global model as follows (10):

$$w_{t+1} \leftarrow \sum_{k=1}^{K} \frac{n_k}{n} w_{t+1}^{k}$$

Here, w_{t+1} represents the resulting global model parameters, with t, K, k, n_k, n, and w^k_{t+1} each representing each round of FL, the number of clients, the index of each client, the amount of data in the client, the total amount of data, and the local model parameters of round t trained in the client, k (if the client, k, does not participate in round t, w^k_{t+1} is equal to w_t). However, recent studies have pointed out that the FedAvg algorithm does not perform well in scenarios with severely heterogeneous data distributions where the data labels as well as the amount of data in each client's label are very different among the clients (11-16). Due to this heterogeneity in data, the client-drift problem in which local client models move away from the global optimal model when updating each client was observed in each client that used the FedAvg algorithm (13). To simulate this data heterogeneity and client-drift, we used the garbage classification dataset containing 2,527 images in 6 classes (cardboard, glass, metal, paper, plastic, and trash) and allocated images of each class according to the Dirichlet distribution—a multivariate continuous probability distribution which is commonly used as a prior distribution in Bayesian statistics and useful to simulate real-world data distributions (17, 18).

By combining FL with the ResNet-18 CNN model, we simulated a waste classification framework and used it to empirically analyze the effects of the following parameters: number of clients, participation ratio, and data heterogeneity. Here, the participation ratio refers to the ratio of the number of active participants in each round of FL to the total number of clients in the system. We decided to focus on the effects of changing data heterogeneity and how to counteract its effects, as it is a big issue associated with FL. We hypothesized that by starting with a pre-trained model and increasing the client participation ratio for FL, we would be able to remedy the negative impact that data heterogeneity had on the model's accuracy. Our results supported our hypothesis, as we could reduce the accuracy degradation by 56% with a higher participation ratio when the number of clients was small. This showcased that the participation ratio played an important role in improving the model's accuracy when heterogeneity existed among small number of clients.

RESULTS

To see how participation ratio affected the performance degradation caused by the heterogeneous data distribution among clients, we kept the other parameters such as the number of rounds of learning at 10 and epochs per client at 5. Then we measured the trained model's accuracies in percent by tweaking the heterogeneity of the data distribution, clients' participation ratio, and the number of clients. During the experiment, we kept the amount of total training data constant, so each client's data size decreased as the number of clients increased. We thought that this setup was fit for observing the effects of the participation ratio since each round of FL had a similar amount of data on average for the same participation ratio regardless of the number of clients.

For training, we used the ResNet-18 model, a CNN that contains 18 neural networking layers and was pre-trained using the ImageNet dataset (7). Then, we fine-tuned the model with a learning rate of 0.000055 using FL. The dataset was sourced from the Kaggle garbage classification dataset containing 2,527 images in 6 classes (**Figure 1**) (17). To simulate heterogeneity similar to real-world situations, we allocated a portion of the samples of each class according to the Dirichlet distribution (18). The heterogeneity of the data distribution was controlled using the concentration parameter β of the Dirichlet distribution. When we tested various β values, we noticed that when β was 10.0, each client had similar amounts of data for all classes and represented the homogeneous data distribution (**Figure 2a**). When β was 0.1, each client had only 1 to 3 data classes and showed enough heterogeneity in data distribution (**Figure 2c**). With this in mind, we picked 0.1, 1.0, and 10.0 as our concentration parameter (β) to represent the heterogeneous, moderate, and homogeneous data distribution (**Figure 2**). To be more accurate to real world environments, we did not regulate the total quantity of data in each client to match each other during data allocation. Consequently, when β decreased, the heterogeneity in the quantity of data in each client

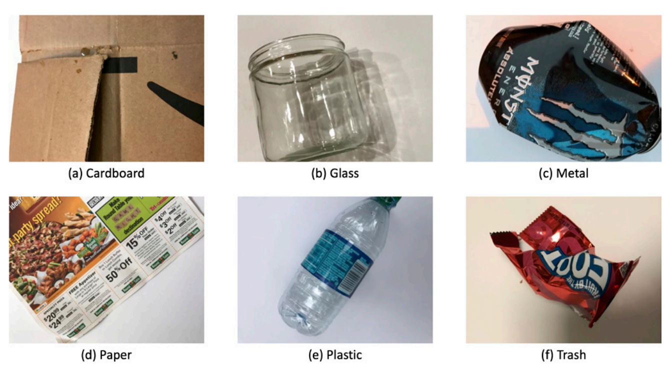


Figure 1. Examples of the six classes in the dataset. Garbage Classification Dataset from Kaggle contains six classes: cardboard (n=403), glass (n=501), metal (n=410), paper (n=594), plastic (n=482) and trash (n=137).

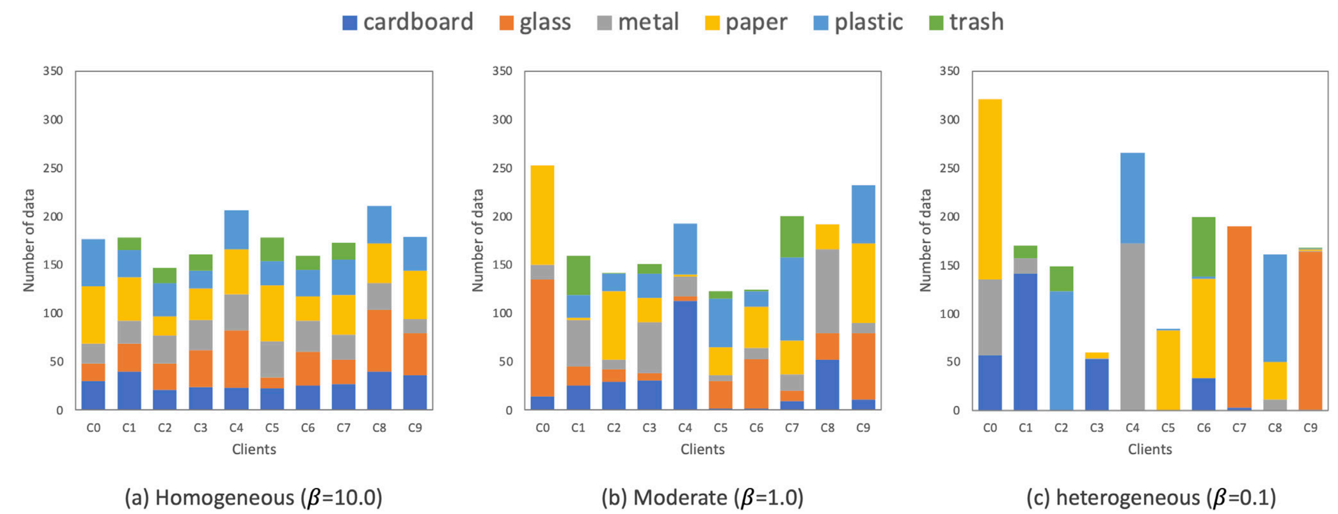


Figure 2. Example data distribution according to the concentration parameter (β): 10.0, 1.0, and 0.1. The x-axis indicates the clients and the y-axis indicates the number of samples for each client. The dataset includes six classes of waste data distributed by the Dirichlet distribution strategy. The heterogeneity was controlled by the concentration parameter (β). **(a)** is an example of a homogeneous distribution where β is 10.0. **(b)** is an example of moderate distribution where β is 1.0. **(c)** is an example of heterogeneous distribution where β is 0.1. Almost all six classes of the data were allocated to each client in homogeneous distribution while only two or three classes of data were allocated to each client in heterogeneous distribution. Note that our allocation algorithm also made the heterogeneity in the amount of data in each client as the concentration parameter decreased.

increased (**Figure 3**). When the number of clients in the system increased, this heterogeneity of quantity also slightly increased, whereas when β was 10.0 the quantity of the data in each client remained almost uniform (**Figure 3a – 3d**). Then, when β was 0.1, we could see the quantity of each clients' data was not uniform (**Figure 3e – 3h)**. Finally, when the number of clients was 15 or 20, we noticed some outliers in the quantity of each clients' data (**Figure 3g, 3h**).

At first, we started the experiment with a heterogeneous data distribution (β=0.1) on 10 clients. As we only had 1,769 images as our training data—70% of our garbage classification dataset—to be partitioned according to the number of clients, we needed to keep the number of clients relatively small to allow each client to have enough data to train their local models. We gradually increased the participation ratio of the clients from 0.2 to 1.0 by a margin of 0.2 and measured the model's accuracy using the test data (**Table 1**). When the participation ratio was 0.2, the accuracy of the trained model was 63.2%. As we increased the participation ratio to 0.4 and 0.6, the accuracy increased to 75.3% and 80.2% respectively. However, when we further increased the ratio to 0.8 and 1.0, the accuracy decreased slightly to 78.8% and then onto 72.4% (**Table 1**). Thus, the accuracy trend showed a concave curve for the participation ratio and showed an even stronger trend with five clients in the system (**Figure 4a, 4b**). With 5 clients, as the participation ratio gradually increased from 0.2 to 0.8, the accuracy of the model increased from 41.3% up to 67.7%, then decreased back down to 59.4% as the participation ratio increased from to 0.8 to 1.0 (**Table 1**).

To calculate the magnitude of the accuracy degradation caused by data heterogeneity, we measured the accuracies of the models trained with the homogeneous data distribution (β=10.0) and the moderately heterogeneous data distribution (β=1.0). When we trained the model with non-heterogeneous data distributions, the accuracies of the trained models were quite high and stable from 79.6% to 90.4% (**Table 1** and

Table 1. The accuracy (%) of the FL models when the numbers of clients in the FL system are 5 and 10.

Number of clients	5					10				
Participation Ratio	**0.2**	**0.4**	**0.6**	**0.8**	**1.0**	**0.2**	**0.4**	**0.6**	**0.8**	**1.0**
heterogeneous (β=0.1)	41.29	61.61	66.36	67.68	59.37	63.19	75.33	80.21	78.76	72.43
moderate (β=1.0)	81.40	87.34	87.34	87.20	85.49	79.55	84.43	84.96	84.70	85.09
homogeneous (β=10.0)	89.31	90.37	88.79	88.79	87.20	85.75	87.60	86.81	87.73	87.99
accuracy degradation	53.77	31.82	25.26	23.78	31.92	26.31	14.01	7.60	10.22	17.68

NOTE: The participation ratio of the clients in each round vary from 0.2 to 1.0. The data distribution was controlled by the concentration parameter (β) of the Dirichlet distribution: 0.1, 1.0 and 10.0. Increasing participation ratio increased the accuracy for the heterogeneous cases and consequently the accuracy for the heterogeneity decreased by 56% (53.77% to 23.78%) at most (when the number of clients is 5).

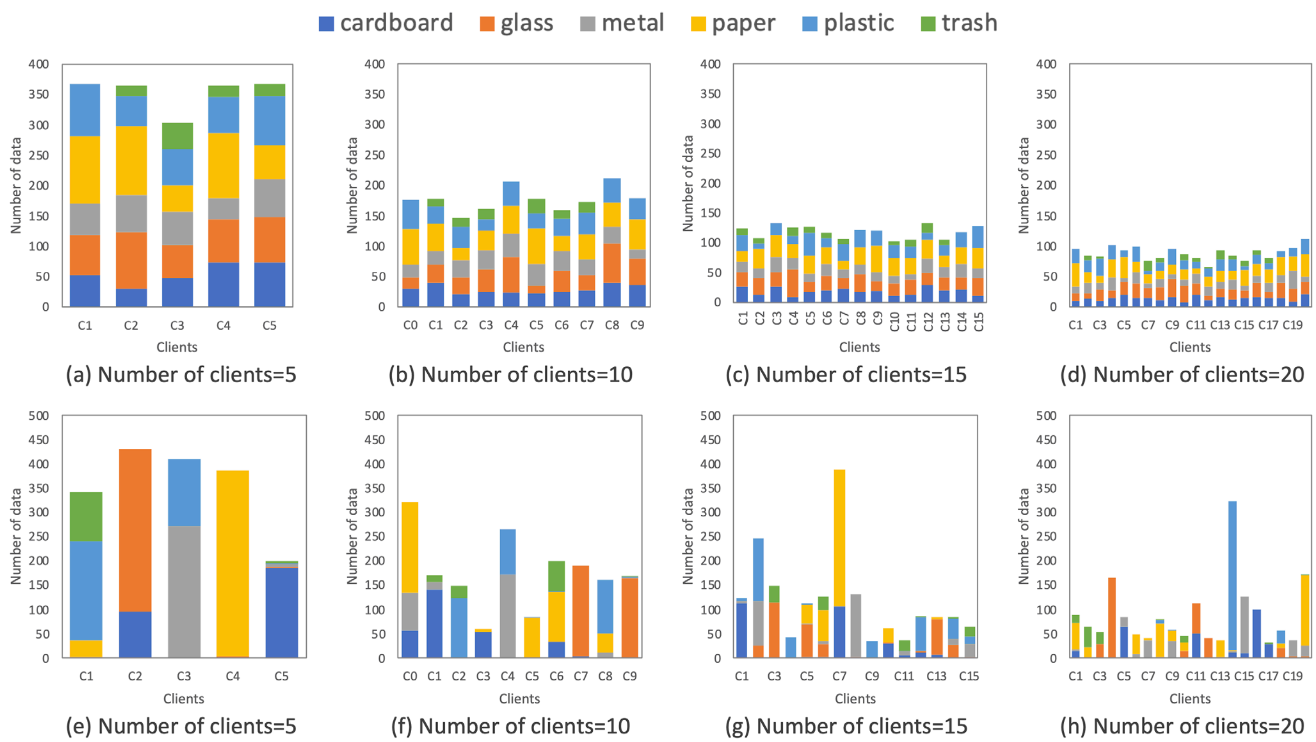


Figure 3. Example data distribution (homogeneous and heterogeneous) according to the numbers of clients. Each graph shows the number of data and their composition in each client for varying numbers of clients: 5, 10, 15, and 20. In the graph, x-axis indicates each client, and y-axis shows the amount of data. **(a-d)** are the example of homogeneous data distribution (β=10.0) and **(e-h)** are the example of heterogeneous data distribution. In a homogeneous distribution, as the number of clients increased, the amount of data decreased uniformly. In heterogeneous distribution, the amount of data also decreased but outliers were observed like client c14 in **(h)**.

Figure 4a, 4b). The accuracy degradation (*Degrad*) caused by the heterogeneity at a certain participation ratio (pr_i)—the ratio of the accuracy difference between the models trained with homogeneous (β=10.0) and heterogeneous (β=0.1) data to the accuracy of the model trained with homogeneous (β=10.0) data—was calculated by the equation as shown below:

$$Degrad_{pr_i} = \frac{accuracy_{pr_i}(\beta = 10.0) - accuracy_{pr_i}(\beta = 0.1)}{accuracy_{pr_i}(\beta = 10.0)} \times 100$$

We observed the maximum accuracy degradation caused by the data heterogeneity, 53.8%, with 5 clients and a participation ratio of 0.2. As the participation ratio increased, the accuracy degradation decreased to 23.8% and then slightly increased to 31.9%. In the case of 10 clients, the magnitude of the accuracy degradation decreased from 26.3% to 7.6% (**Table 1** and **Figure 5**).

On the other hand, when we increased the number of clients to 20, the participation ratio did not affect the accuracy of the model trained with the heterogeneous data distribution (**Figure 4d**). The accuracy of the model at the participation ratio, 0.2 was 56.2% and when we increased the ratio to 0.4, 0.6, 0.8, and 1.0, the accuracy was just slightly fluctuated to 61.5%, 58.3%, 62.5%, and then 60.4% respectively (**Table 2**). This trend was similar when the model was run with 15 clients (**Figure 4c**). The accuracies of the models trained with non-heterogeneous data distributions were high and stable from 80.6% to 86.9%, and the accuracies of the models trained

Table 2. The accuracy (%) of the FL models when the numbers of clients in the FL system are 15 and 20.

Number of clients	15					20				
Participation Ratio	**0.2**	**0.4**	**0.6**	**0.8**	**1.0**	**0.2**	**0.4**	**0.6**	**0.8**	**1.0**
heterogeneous (β=0.1)	62.27	64.51	68.07	65.30	67.68	56.20	61.48	58.31	62.53	60.42
moderate (β=1.0)	84.96	83.38	84.96	83.64	84.17	80.61	81.93	82.72	82.45	81.40
homogeneous (β=10.0)	86.94	85.22	85.22	85.75	86.15	83.64	84.17	83.51	83.77	83.38
accuracy degradation	28.38	24.30	20.12	23.85	21.44	32.81	26.96	30.18	25.36	27.54

NOTE: In these cases, increasing participation ratio did not show meaningful accuracy increment when the data distribution was heterogeneous. Consequently, there was no significant accuracy degradation decrease along with the participation ratio.

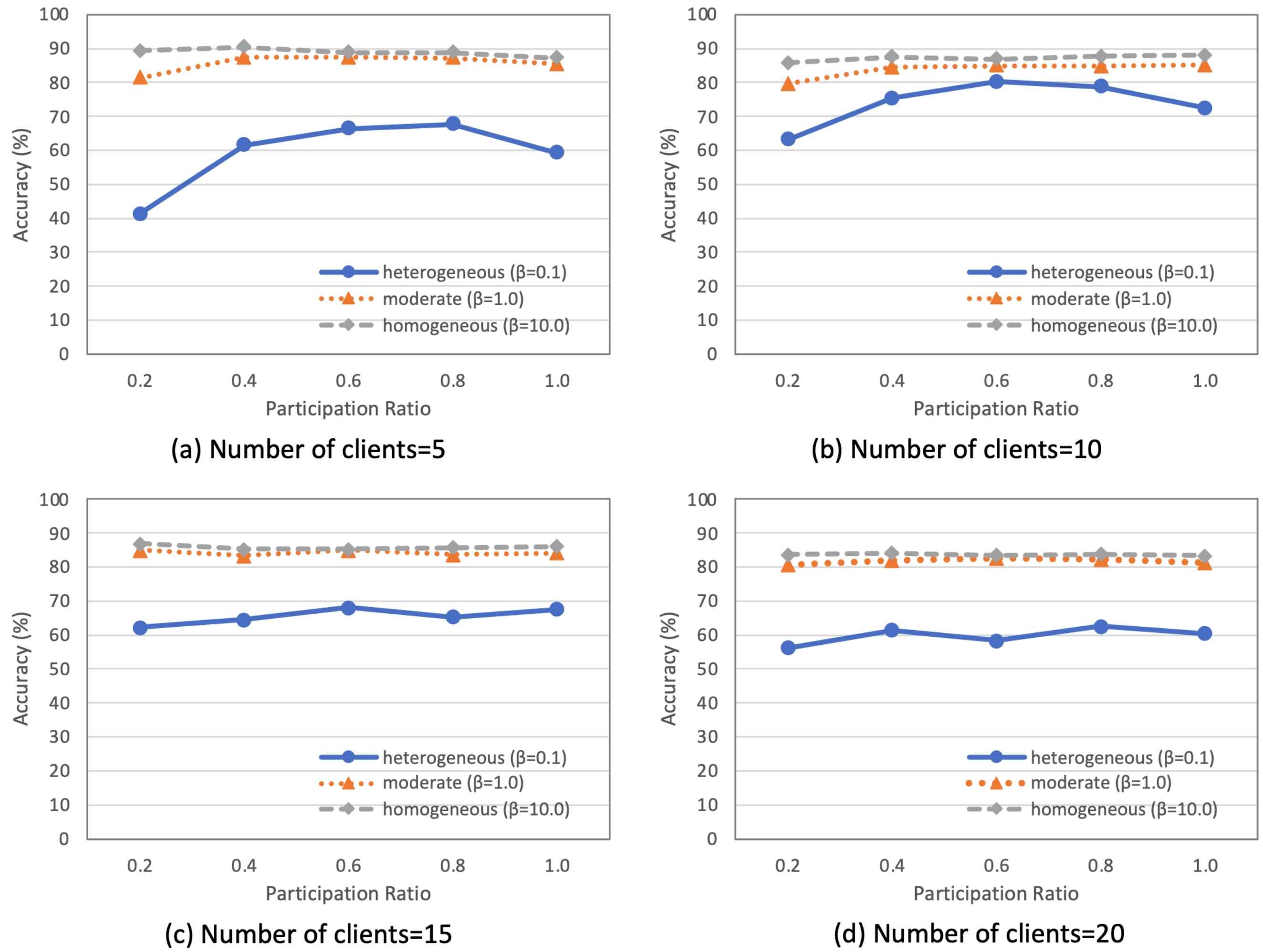


Figure 4. Accuracy of the waste classification model. Each graph shows the accuracy changes by the participation ratio of the clients in each round. (**a-d**) are when the numbers of the clients in each FL round are 5, 10, 15, and 20 respectively. In each graph, the x-axis is the participation ratio of the clients, and the y-axis is the model accuracy (%) evaluated by the test data. The graph shows the fact that the heterogeneity in data distribution caused the accuracy degradation in each model. As we expected, the accuracy of the models trained with 5 and 10 clients in heterogeneous distribution increased as the participation ratio increased, but the accuracy of the model trained with 15 and 20 clients was not affected by the participation ratio.

with a heterogeneous data distribution were relatively low and stable, ranging from 62.3% to 68.1% (number of clients 15) and from 56.2% to 62.5% (number of clients 20) (**Table 2**). The corresponding graph of the accuracy shows this trend more explicitly as the curves of non-heterogeneous data (β=1.0, β=10.0) had higher accuracies, while the curves of heterogeneous had lower accuracies. All curves showcased negligible changes in accuracy due to participation ratio (**Figure 4c, 4d**). Consequently, when we looked at the accuracy degradation trend graph, we could not see meaningful changes by the participation ratio as the curves were relatively flat and just slightly fluctuated compared to the accuracy degradation trend when the number of clients were 5 and 10 (**Figure 5**).

To investigate why increasing the number of clients resulted in decreasing the effectiveness of the participation ratio, we designed an experiment measuring the accuracies as the number of clients changed. In our experiment, we kept the participation ratio at 1.0, meaning that all clients participated in the learning process to eliminate the effect of the participation ratio. Increasing the number of clients tended to generally decrease the accuracy of the models trained with all data distributions except one case: the model trained with heterogeneous data with 5 clients (**Figure 6a**).

Finally, we wanted to see how strongly the accuracy was affected by the participation ratio. For that, we calculated the standard deviation of the accuracy changes by the participation ratio in each number of clients and concentration parameters (β) (**Figure 6b**). Here, the larger standard deviation means the participation ratio affected the accuracy more. From the calculation, we noticed that the largest standard deviation was the case of five clients with heterogeneous data. The standard deviation with the heterogeneous data was decreased as the number of clients increased but still larger than the other data distributions: moderate and homogeneous (**Figure 6b**).

DISCUSSION

The experiments gave insight on the different variables that contributed to the model's performance and their relationships to one another. We tested the change in performance with data heterogeneity, participation ratio, varying client counts, and corresponding amount of data in each client.

Regarding the data heterogeneity, the difference between a homogeneous data distribution and a heterogeneous data distribution was whether each client's local dataset was complete—containing all or almost all classes of the data—or not. A client belonging to a homogeneous data distribution system had a higher chance that its local training data was complete (**Figure 2a**). However, a client in a heterogeneous data distribution had only about one or two out of the six classes of the whole training data in the local training data (**Figure 2c**). This means that the clients with heterogeneous data had more chance of the client-drift which degraded the accuracy of the global model. This can explain why the accuracy of the model trained with heterogeneous data was lower than other cases (**Figure 4**).

To see the effects of participation ratio on heterogeneous data, we need to consider the characteristics of the whole sum of all participants' data in each training round of FL. In situations where the participation ratio was high, even if the data in each client was incomplete, the participants' data as a whole in each round was almost complete. Thus, the models trained with higher participation ratio showed higher accuracies than the lower participation ratio given that there was a small number of clients (**Figure 4a, 4b**). In the aspect of the accuracy degradation by the data heterogeneity, the accuracy degradation decreased when the participation ratio increased (#clients=5 and #clients=10, **Figure 5**).

However, we also noticed that the model's accuracy was quite low and stable even though the participation ratio increased when the number of clients increased to 15 and 20 (**Figure 4c, 4d**). As a result, the accuracy degradation by the heterogeneous data was not affected by the participation ratio (#clients=15 and #clients=20, **Figure 5**). There are three possible reasons for this. The first is related with the amount of data in each client. Since the total amount of training data was fixed (1,769) and distributed to each client, having more clients simply means that each client had less data for training (**Figure 3a – 3d**). This was best represented by the measurements when the data was homogeneous (β=10.0). Since all six classes of the data were evenly distributed, the very gradual accuracy decrease was a direct result of having

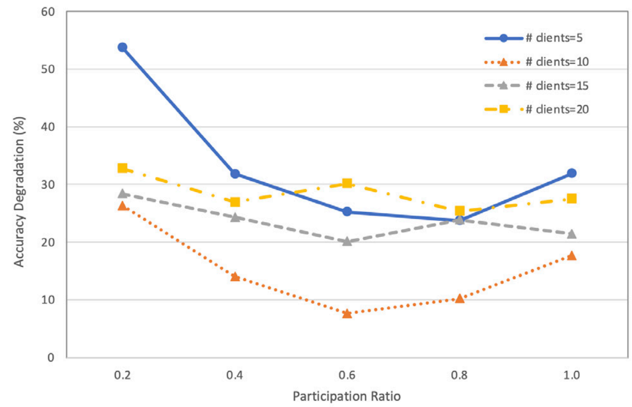
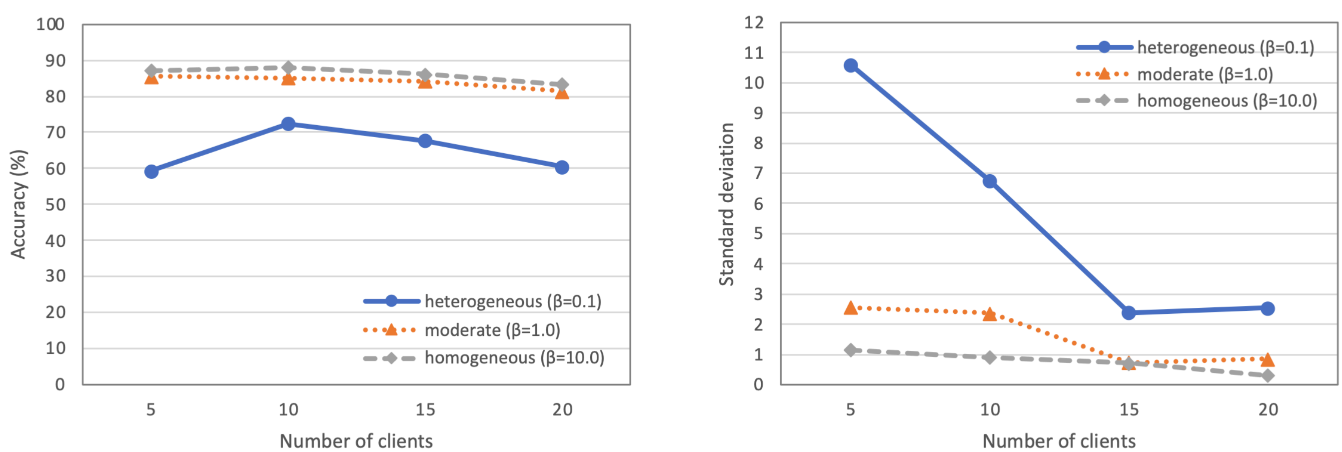


Figure 5. Accuracy degradation by heterogeneity. Each graph shows the accuracy degradation by the heterogeneity of the data distribution against the participation ratio of clients in each round. The participation ratio increments meaningfully recovered the accuracy degradation when the number of clients were 5 and 10, but there was no recovery of the accuracy degradation when the number of clients were 15 and 20.



(a) Accuracy (participation ratio=1.0) (b) Standard deviation of accuracies

Figure 6. The accuracy (participation ratio=1) and the standard deviation of accuracies of various participation ratios on the different number of clients. The standard deviation was calculated with the accuracies of the varying participation ratio: 0.2, 0.4, 0.6, 0.8, and 1.0. **(a)** shows the accuracy trend of the models trained with different data distributions (when participation ratio = 1.0) and **(b)** shows the effect of the participation ratio to the accuracy of the models when the number of clients increased. The participation ratio had huge effect on the accuracy of the model trained with 5 and 10 clients in heterogeneous data distribution.

less data in each client (homogeneous, **Figure 6a**). This partially affected the rapid accuracy decrease in the models trained with heterogeneous data as the number of clients increased from 10 to 20 compared to the accuracy decrease when the data distribution was homogeneous or moderate (heterogeneous, **Figure 6a**).

The second reason lies within the heterogeneity simulation we used to simulate the realistic distribution. Our heterogeneous distribution also included some outliers in the amount of data in each client (**Figure 3e – 3h**). The clients corresponding to those outliers experienced more 'client-drift' by the biased data and caused the further global model accuracy degradation and partially affected the decreased accuracy by increasing the number of clients (heterogeneous, **Figure 6a**).

Finally, we should look at the effect that the client participation ratio had on the model's accuracy as the number of clients increased. This effect was measured using the standard deviation of the accuracies of the models with varying participation ratio. When the number of clients increased, the effect of the participation ratio on the accuracy decreased (**Figure 6b**). For example, when the participation ratio was 0.2, the accuracy of the model, which was trained with 20 clients, was much higher than the model with 5 clients. With the same participation ratio, having a larger number of clients implies a larger number of participants, meaning that there was a higher chance of having all classes in the data of the participants. In our heterogeneous case, with five clients, only one participant with maximum three classes participated in each round. On the other hand, with 20 clients, 4 participants with maximum 3 classes participated, which means it is more likely for each round of FL to have all 6 classes of data (**Figure 3e, 3h**). Consequently, in each round of FL, the effect of the client-drift of each participant on the global model counterbalanced each other. The sum of the participants' data started to resemble the total training dataset across all values for the participation ratio starting from 0.2, and this negated the effect that the participation ratio had on the accuracy. This was also highly related to the fact that our data had only six classes.

Contrary to our results, normally in the real world, having more clients means that the model can access more data because the total amount of data itself is not limited, and consequently the model's accuracy should increase. In the case of waste classification, the probability of the clients having similar waste images as the types of waste becomes limited would increase. This effect can be simulated by allowing duplicate data allocation among the clients. We decided to leave this to our future works. As our experiments showed, the data heterogeneity had a big impact on the accuracy of the FL framework. To counteract the impact of heterogeneity, we can also think of a way to reduce the heterogeneity itself in each client by sending the clients a small but complete sample dataset from the server. In this case, the server's dataset should be a public dataset not to harm the privacy characteristics of FL. Regarding the experiments, when the participation ratio was as low as 0.2, the number of clients participating in FL round was small, for example, one for 5 clients or four for 20 clients. This means the measured accuracy was highly dependent on the clients that participated in each round and our measurement was not highly confident. To overcome this kind of confidence issue, we need to measure the accuracies multiple times and present the accuracies with variations in our future works.

We started our research to solve the problem of the lack of data in building a waste classification system. In nature, this application assumes many clients and low participation ratio. But the number of classes in the data is relatively small (six in our case) and thus, it is more probable that each client has all categories of data despite the size of the data. This means that the data distribution is quite homogeneous. In this paper, however, we showed that the participation ratio played a key role in improving the accuracy in heterogeneous data distribution with small number of clients. The work we conducted here is useful for informing future work on waste management but can also be applied to other fields. One way our study could inform other important scenarios beyond waste management could be the application of using FL to build a model for screening patients with medical data among big hospitals because the patient data should be kept privately in each hospital. In this case, the number of clients is limited while the number of categories is big. Consequently, our findings of participation ratio on the effect of accuracy degradation by the heterogeneous data distribution can be applied.

MATERIALS AND METHODS

For our experiment, we used Python 3.8.10 and the pre-trained ResNet-18 of PyTorch 1.13.1 as our starting model in the server. We then used the FL algorithm to fine-tune the model. Instead of building a distributed system for the FL framework, we simulated an FL framework on a single server in Google Colab Pro+. For the training data, we used the garbage classification dataset from Kaggle (17). It consists of six classes and each class has a different number of samples (**Figure 1)**: cardboard (n=403), glass (n=501), metal (n=410), paper (n=594), plastic (n=482), trash (n=137). We divided the dataset into 30% test data (n=758) and 70% training data (n=1,769). For the FL simulation, the training data was then distributed to simulated clients using the Dirichlet distribution in order to simulate data heterogeneity. We used the Dirichlet function in NumPy library in Python to draw samples from the Dirichlet distribution for each class and assigned the data proportion to the distribution. During the assignment, we guaranteed that none of the clients had less than 30 samples in total to allow the local training to have a notable effect on the global model. The concentration parameter (β) of the function determines the level of heterogeneity of the resulting sample distribution. If the β is small, the data becomes more heterogeneous, and vice versa. As our concentration parameter (β), we chose 0.1 for the heterogeneous distribution, 1.0 for the moderate distribution, and 10.0 for the homogeneous distribution (**Figure 2**).

Because of the limited amount of test data, we chose 5, 10, 15, and 20 clients for the test and varied the client participation ratio from 0.2 to 1.0 in increments of 0.2 and measured the model's accuracy after 10 rounds of FL. In each client, as our base model was the pre-trained version of ResNet-18 and we fine-tuned the model with our garbage dataset, we fixed the learning rate to a very small number of 0.000055 and trained with local epoch of 5.

To gauge the effects that the other variables have on the performance degradation caused by data heterogeneity, we calculated the accuracy degradation ($Degrad_{pri}$) for each participation ratio (pr_i) and the number of clients (**Table 1** and **Table 2**). To check how increasing the number of clients diminishes the effect of the participation ratio, we calculated the standard deviation of the different accuracy values according to the participation ratio for each number of clients (**Figure 6b**).

Received: June 28, 2022
Accepted: June 15, 2023
Published: July 18, 2023

REFERENCES

1. "National Overview: Facts and Figures on Materials, Wastes and Recycling." *US EPA*. www.epa.gov/facts-and-figures-about-materials-waste-and-recycling/national-overview-facts-and-figures-materials. Accessed 30 Oct. 2022.
2. "How the Recycling Center Works." *About the Recycling Center - Montgomery County, Maryland*, www.montgomerycountymd.gov/SWS/facilities/rc/how.html. Accessed 20 May 2023.
3. Casey, Tina. "Why Labor Shortages Could Stall Any Recycling Progress." *TriplePundit*, 31 Jan. 2022, www.triplepundit.com/story/2022/labor-shortage-recycling/735681. Accessed 20 May 2023.
4. Adedeji, Olugboja, and Zenghui Wang. "Intelligent Waste Classification System Using Deep Learning Convolutional Neural Network." *Procedia Manufacturing,* vol. 35, 2019, pp. 607-612, doi:10.1016/j.promfg.2019.05.086.
5. Vo, Anh H., *et al*. "A Novel Framework for Trash Classification Using Deep Transfer Learning." *IEEE Access,* vol. 7, Dec. 2019, pp. 178631-178639, doi:10.1109/ACCESS.2019.2959033.
6. Ahmed, Lulwa, *et al*. "Active Learning Based Federated Learning for Waste and Natural Disaster Image Classification." *IEEE Access,* vol. 8, Nov. 2020, pp. 208518-208531, doi:10.1109/ACCESS.2020.3038676.
7. He, Kaiming, *et al*. "Deep Residual Learning for Image Recognition." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770-778.
8. Xie, Saining, *et al*. "Aggregated Residual Transformations for Deep Neural Networks." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 1492-1500.
9. Deng, Jia, *et al*. "ImageNet: A Large-Scale Hierarchical Image Database." *2009 IEEE conference on computer vision and pattern recognition*, Jun. 2009, pp. 248-255, doi:10.1109/CVPR.2009.5206848.
10. McMahan, Brendan, *et al*. "Communication-Efficient Learning of Deep Networks from Decentralized Data." *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, vol. 54, 2017, pp. 1273-1282.
11. Sahu, Anit Kumar, *et al*. "On the Convergence of Federated Optimization in Heterogeneous Networks." *arXiv preprint arXiv:1812.06127v1*, 2018, doi:10.48550/arXiv.1907.02189.
12. Li, Xiang, *et al*. "On the Convergence of FedAvg on Non-IID Data." *arXiv preprint arXiv:1907.02189v4, 2019*, doi:10.48550/arXiv.1907.02189.
13. Karimireddy, Sai Praneeth, *et al*. "SCAFFOLD: Stochastic Controlled Averaging for Federated Learning." *Proceedings of the 37th International Conference on Machine Learning*, vol. 119, Jul. 2020, pp. 5132-5143.
14. Hsu, Tzu-Ming Harry, *et al*. "Measuring the Effects of Non-Identical Data Distribution for Federated Visual Classification." *arXiv preprint arXiv:1909.06335v1,* 2019, doi:10.48550/arXiv.1909.06335.
15. Wang, Jianyu, *et al*. "Tackling the Objective Inconsistency Problem in Heterogeneous Federated Optimization." *Advances in Neural Information Processing Systems*, vol. 33, 2020, pp. 7611-7623.
16. Reddi, Sashank, *et al*. "Adaptive Federated Optimization." *arXiv preprint arXiv:2003.00295v5*, 2020, doi:10.48550/arXiv.2003.00295.
17. "Garbage Classification." *Kaggle*, 24 Nov. 2018, www.kaggle.com/ds/81794. Accessed 30 Oct. 2022.
18. Li, Qinbin, *et al*. "Federated Learning on Non-IID Data Silos: An Experimental Study." *2022 IEEE 38th International Conference on Data Engineering (ICDE)*, May 2022, doi:10.1109/ICDE53745.2022.00077.