

# A Comparative Study of Federated and Centralised Learning for Waste Classification with Non-IID Data

Original Scientific Paper

## Qian Wang

Universiti Teknologi MARA,  
Faculty of Computer and Mathematical Sciences  
Shah Alam, Selangor, Malaysia  
2022174061@student.uitm.edu.my  
wqian8603@gmail.com

## Lei Wang

Universiti Teknologi MARA,  
Faculty of Computer and Mathematical Sciences  
Shah Alam, Selangor, Malaysia  
1105135335@qq.com,  
2024470582@student.uitm.edu.my

\*Corresponding author

## Shafaf Ibrahim

Universiti Teknologi MARA,  
Faculty of Computer and Mathematical Sciences  
Shah Alam, Selangor, Malaysia  
shafaf2429@uitm.edu.my

## Zainura Idrus\*

Universiti Teknologi MARA,  
Faculty of Computer and Mathematical Sciences  
Shah Alam, Selangor, Malaysia  
zainura@fskm.uitm.edu.my

**Abstract** – Accurate waste classification is essential for sustainable environmental management, as traditional manual approaches are time-consuming, labour-intensive, and prone to human error. Deep Learning (DL) has achieved remarkable progress in image-based classification, but its dependence on large, labelled datasets and centralised training raises concerns about data privacy and scalability. Federated Learning (FL) provides a privacy-preserving alternative by enabling model training across decentralised devices without sharing raw data. However, applying FL to waste image classification remains challenging due to the non-independent and identically distributed (non-IID) nature of client data, caused by variations in environment, lighting, and user habits. To address this, we propose a privacy-preserving and adaptive FL framework tailored for waste image classification under heterogeneous data distributions. Five Convolutional Neural Network (CNN) architectures—ResNet-18, ResNet-50, GoogLeNet, DenseNet-121, and VGG-19—were systematically compared under both centralised DL and FL settings. Experimental results show that GoogLeNet achieves the highest accuracy, reaching 80.45% under non-IID FL conditions, outperforming centralised DL by up to 9.8% in specific configurations. These findings demonstrate the effectiveness of FL in improving generalisation and robustness while preserving privacy, providing practical insights for developing scalable, intelligent waste management systems in real-world, diverse environments.

**Keywords:** Federated Learning (FL), non-IID data; Deep Learning (DL), Convolutional Neural Networks (CNNs), waste classification, adaptive aggregation, data heterogeneity, privacy-preserving intelligence, smart waste management

Received: August 9, 2025; Received in revised form: December 8, 2025; Accepted: December 26, 2025

## 1. INTRODUCTION

As global waste production surges, efficient and accurate waste classification has become a cornerstone of sustainable environmental management. Manual sorting, still common in many regions, is time-consuming, error-prone, and physically demanding. Automated systems powered by Machine Learning (ML) and Deep Learning (DL) have emerged as promising solutions, offering fast and reliable image-based waste recognition. However, traditional DL approaches require large, centralised datasets, which can pose privacy risks and

scalability challenges, especially when handling sensitive or distributed data sources [1].

Federated Learning (FL) provides a compelling alternative by enabling collaborative model training across multiple devices without sharing raw data. This approach is particularly valuable for waste classification, where data is often collected from diverse sources—such as households, recycling centres, or smart bins—resulting in non-independent and identically distributed (non-IID) data distributions. Non-IID data presents unique challenges, as client devices may have entirely different waste categories or imbalanced datasets, which can complicate

model convergence and performance [2], [3]. While prior studies, such as Liu et al. [4] and Yang et al. [2], have applied DL to waste classification with high accuracy (e.g., 93–97%), their reliance on centralised data collection limits their applicability in privacy-sensitive scenarios. Li et al. [3] demonstrated FL's potential for waste classification, achieving robust results without data sharing, but a systematic comparison of FL and DL under non-IID conditions remains underexplored.

This study addresses this gap by developing a privacy-preserving FL framework for waste image classification, specifically designed to handle non-IID data. We compared the performance of five widely used CNN architectures—ResNet-18, ResNet-50, GoogLeNet, DenseNet-121, and VGG-19—under both centralised DL and FL settings. Our experiments tested various data partitioning strategies, including IID, Dirichlet-based non-IID ( $\alpha = 0.1, 0.01$ ), and sharding-based non-IID ( $s = 4, 2$ ), to simulate real-world heterogeneity. Our key contributions are:

1. A detailed comparative analysis of federated and centralised deep learning, revealing FL's unique robustness to non-IID waste data with up to 5% accuracy gains over centralised methods.
2. A novel FL framework developed to underpin this comparative study, enabling evaluations across diverse data distributions and highlighting accuracy, stability, and sensitivity to non-IID conditions.
3. Evaluation of five CNN architectures providing targeted guidance for model selection and hyperparameter tuning in real-world waste management systems.

By addressing the challenges of non-IID data and privacy concerns, this study paves the way for scalable, privacy-compliant smart waste management solutions.

## 2. RELATED WORK

Recent advances in waste classification have leveraged both traditional ML and DL techniques, with varying degrees of success, as shown in Table 1. Yang et al. [5] introduced the TrashNet dataset, containing 2,527 images across six waste categories, achieving 63% accuracy using a support vector machine (SVM). Satvilkar [6] reported a 62.61% accuracy rate with a random forest classifier, later improving to 70.1% with XGBoost. Costa et al. [7] applied K-nearest neighbours (KNN), reaching 88% accuracy. While these traditional ML methods are computationally lightweight, they often require complex preprocessing pipelines, which can lead to storage overhead and deployment challenges.

DL approaches have significantly outperformed traditional ML in waste classification. Rabano et al. [8] used MobileNet on the TrashNet dataset, achieving 87.2% accuracy on a mobile device. Adedeji et al. [9] combined ResNet-50 with an SVM classifier, achieving 87%. More recently, Aral et al. [10] fine-tuned DenseNet-121 and Incep-

tionResNetV2, achieving 95% and 94% accuracy, respectively. Ozkaya et al. [11] achieved a remarkable 97.86% accuracy using a GoogLeNet-SVM combination on TrashNet. Beyond public datasets, Mittal et al. [12] developed the GINI dataset (2,561 images) and proposed GarbNet, achieving 87.69% accuracy. Yang et al. [2] introduced GarbageNet, which achieved 96.96% accuracy on the Huawei Cloud Garbage Classification Challenge dataset. Fu et al. [13] reported 93.47% and 92.62% accuracy rates, using EfficientNet and an enhanced MobileNetV3.

Despite these advances, most DL-based studies rely on centralised training, which raises privacy concerns when data is collected from distributed sources like smart bins or user devices. Federated Learning has emerged as a solution to address these concerns [14]. McMahan et al. [15] introduced the Federated Averaging (FedAvg) algorithm, demonstrating its efficiency in training deep networks across decentralised devices. Hard et al. [16] applied FedAvg to mobile keyboard prediction, enhancing privacy without compromising performance. Bonawitz et al. [17] outlined scalable FL system designs that address challenges such as communication efficiency and client availability. In the domain of waste classification, Li et al. [18] combined FL with graph convolutional networks, achieving high accuracy while preserving data privacy. Recent studies have explored FL on non-IID data, but mostly for general image classification [18][19]; for waste management, deep-learning approaches that preserve privacy have been proposed [20][21].

Recent works have increasingly applied FL to waste management, focusing on non-IID data challenges in distributed environments. While centralised approaches can still deliver strong results in controlled settings. For instance, Ahmad et al. (2025) [22] proposed a highly memory-efficient CNN incorporating Global Average Pooling that achieved 92.5% accuracy on a custom four-class waste dataset (aluminum, cardboard, plastic, and glass), using just 495,620 parameters and 1.27 GFLOPs, dramatically outperforming bulkier models like VGG16 and ResNet50 in both efficiency and several performance metrics, these methods inherently lack the privacy protections that FL provides in decentralized, multi-stakeholder scenarios. Kumar et al. (2024) [21] proposed a federated learning system for IoT-based waste management using models such as ResNet50, VGG16, InceptionV3, and EfficientNet, along with FedAvg, FedAMP, FedAdam, and FedMA. On a custom 15,150-image dataset spanning 12 waste categories, FedAdam achieved the highest accuracy at 77.56%. The approach highlights privacy preservation in decentralised smart bins but does not explicitly address extreme non-IID conditions such as pathological data distributions. Shami et al. (2025) [23] introduced federated continual learning (FCL) for vision-based plastic classification in recycling, using ResNet-18 on the custom DWRL dataset (6 plastic types). Their approach combined FL strategies (FedAvg, FedOpt, FedProx) with continual learning techniques (EWC, Replay), yielding 83.68% overall accuracy

and robustness to non-IID data across multiple plants; however, it was constrained by dataset size and simulated environments, highlighting scalability issues for real-world deployments. Vignesh and Subashini (2025) [24] developed a multilevel thermoplastic waste segregation system using AHGSO (Adaptive Hybrid Grey Wolf Spider Monkey Optimisation) with FL on a custom dataset, achieving improved classification accuracy (up to 98% in some classes) under non-IID conditions; the work's strength lies in hybrid optimisation for heterogeneous data, but it lacks comparison with centralised DL baselines. Zhang et al. (2024) [25] presented FedScrap, a layer-wise personalised FL framework for scrap detection (waste metal classification), tested on industrial datasets with non-IID distributions, reaching high accuracy (up to 95%) via personalised aggregation; its limitation is domain specificity to scrap, with less emphasis on privacy metrics. Alharithi and Alzahrani (2024) [26] applied federated LSTM models to simulated non-IID data for environmental sustainability and waste management, achieving 85–90% classification accuracy, but did not evaluate on real-world waste datasets. Similarly, Nooh (2025) [27] proposed an IoT-based FL approach with edge computing for low-carbon environmental monitoring, achieving up to 92% accuracy and 20–30% energy reduction, though its focus on monitoring limits direct applicability to waste sorting.

Compared to prior work, our study stands out by systematically comparing FL and DL under varying non-IID conditions, using a diverse set of CNN architectures. Unlike studies conducted by Yang et al. [2] and Ozkaya et al. [11], which achieved high accuracy but relied on centralised data, our FL framework ensures data privacy while maintaining competitive performance. Additionally, we address the limitations of earlier FL studies [15], [18] by evaluating extreme non-IID scenarios (e.g.,  $\alpha = 0.01$ ,  $s = 2$ ), providing deeper insights into model robustness and practical deployment.

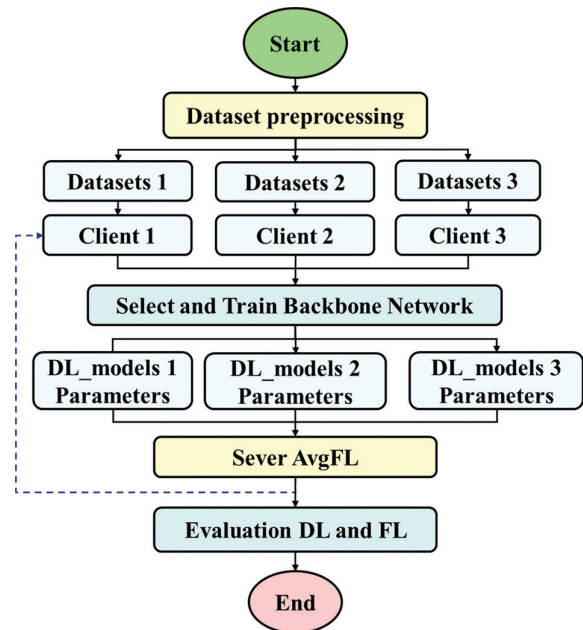
**Table 1.** Related works

Reference	Dataset	Method	Accuracy
Yang et al.	TrashNet	SVM	63%
Satvilkar et al.	TrashNet	Random Forest	62.61%
Satvilkar et al.	TrashNet	XGBoost	70.1%
Costa et al.	TrashNet	KNN	88.0%
Rabano et al.	TrashNet	MobileNet	87.2%
Ruiz et al.	TrashNet	Inception-ResNet	88.6%
Adedeji et al.	TrashNet	ResNet-50 + SVM	87%
Aral et al.	TrashNet	DenseNet121 / InceptionResNetV2	95% / 94%
Ozkaya et al.	TrashNet	GoogleNet + SVM	97.86%
Mittal et al.	GINI	GarbNet	87.69%
J. Yang et al.	Huawei Cloud Dataset	GarbageNet	96.96%
Fu et al.	Not specified	GNet (Improved MobileNetV3)	92.62%
McMahan et al.	Distributed devices	FedAvg	-
Hard et al.	Google Keyboard	FedAvg	-
Ahmad et al.	Customer	Efficient CNN	92.5%

### 3. METHODOLOGY

#### 3.1. FRAMEWORK

Our proposed FL framework, illustrated in Fig. 1 and Algorithm 1, enables decentralised waste classification while preserving data privacy. The dataset is pre-processed and distributed across multiple clients, each of which trains a local model (e.g., ResNet-18, GoogLeNet) on its unique data subset. Local model parameters are sent to a central server, which aggregates them using the Federated Averaging (FedAvg) algorithm. Algorithm 1 Federated Learning Framework for Waste Classification. The aggregated global model is then compared to centralised DL models to evaluate performance under varying data distributions.



**Fig. 1.** Federated Learning Framework for Waste Classification (The framework distributes datasets across clients, who train local models independently.) The central server aggregates model parameters using FedAvg and redistributes the global model for iterative updates.)

The framework supports three data partitioning strategies: IID (uniform random sampling), Dirichlet-based non-IID (controlled by  $\alpha = 0.1, 0.01$ ), and sharding-based non-IID (controlled by  $s = 4, 2$ ). These strategies simulate real-world scenarios where client data varies significantly, ensuring robustness across heterogeneous environments.

#### Algorithm 1: Federated Learning Framework for Waste Classification

**Input:** Number of clients  $K$ , communication rounds  $T$ , learning rate  $\eta$ , local epochs  $E$ , data partitioning strategy (IID/Dirichlet/Sharding with  $\alpha$  or  $s$ ), pre-trained models (ResNet-18, etc.), non-IID parameter ( $\alpha=0.1/0.01$  or  $s=4/2$ )

**Output:** Global model  $W_{global}$

**Server executes:**

```

Initialise global model  $W_0$ 
for each round  $t = 1$  to  $T$ :
    Select a subset of clients  $S_t$  (e.g., 6 out of 10)
    for each client  $k = 1$  to  $K$  in parallel:
        // Local training
        Receive global model  $W_t$ 
        Partition local data  $D_k$  using strategy
        (IID/Dirichlet/Sharding)
        for each local epoch  $e = 1$  to  $E$ :
            Compute loss  $L$  on mini batch from  $D_k$ 
            Update local model:
             $w_k \leftarrow w_k - \eta \nabla L(w_k; D_k)$ 
            Send local update  $\Delta w_k = w_k - W_t$  to server
        end
    end
end
end

```

**Aggregate:**  $W_{t+1} = W_t + (1/|S_t|) \sum \Delta w_k$  // FedAvg  
 return  $W_{global}$

**3.2. DATASET PREPROCESSING**

This study employed the TrashNet dataset, compiled by Mindy Yang and Gary Thung from Stanford University [19], comprising 2,400 images (512 × 384 pixels) across six waste categories: paper, glass, plastic, metal, cardboard, and trash. Each category contains 400–500 images, except for the "trash" class (~100 images). Images were captured against a consistent white background to minimise noise.

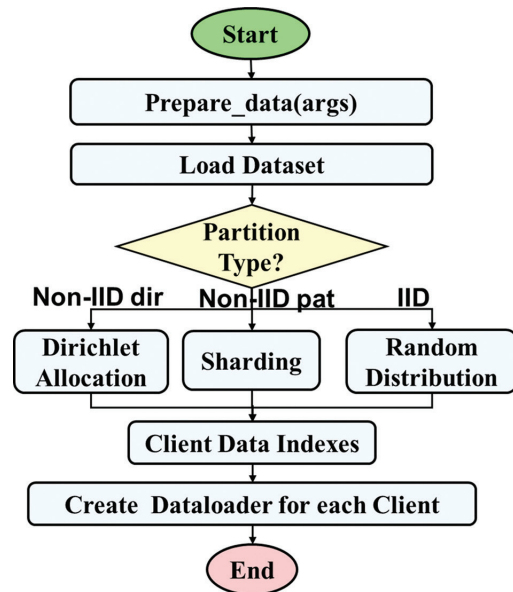


**Fig. 2.** Sample Images from TrashNet Dataset (a) Paper, (b) Glass, (c) Plastic, (d) Metal. Images are uniformly captured against a white background for consistency)

To comprehensively evaluate the robustness of our method under varying levels of data heterogeneity, we adopted three s used data partitioning strategies in federated learning: random split, label-based sharding, and Dirichlet distribution-based partitioning.

In the IID setting, the data were randomly shuffled and evenly distributed across clients to simulate the same distribution. For the extreme non-IID scenario, we employed the label-based sharding strategy proposed by McMahan et al. [15], where the dataset was sorted by class label and partitioned into small shards according to a heterogeneity control parameter  $s \in \{4, 2\}$ . Each client was assigned a fixed number of shards, so that the client holds data from only a limited subset of classes. When  $\alpha$  is smaller, data across clients becomes more uneven. Following Lin et al. [14], we control for heterogeneity by assigning class proportions to each client using a Dirichlet distribution with  $\alpha = 0.1$  or  $0.01$ . The lower the  $\alpha$ , the more imbalanced the class distribution for each client.

This approach lets us test our method across a spectrum—from balanced (IID) to highly skewed (non-IID) data, which more closely reflects real-world federated learning conditions.



**Fig. 3.** Dataset preprocessing

**3.3. DATA PARTITIONING**

In real-world FL, data is naturally non-IID due to variations in waste generation across clients (e.g., households vs. recycling centres). To simulate this, we implemented two non-IID partitioning methods:

Dirichlet-based (LDA): Controlled by  $\alpha$ , where smaller values (e.g., 0.01) result in highly imbalanced class distributions.

Sharding-based: Data is sorted by class and split into shards, with each client receiving  $s$  shards ( $s = 4, 2$ ).  $s$  values increase heterogeneity.

These settings ( $\alpha = 0.1, 0.01$ ;  $s = 4, 2$ ) were chosen to test extreme non-IID conditions, extending beyond typical values ( $\alpha = 0.3-1.0, s = 10$ ) used in prior work [14], [19].

### 3.4. DL MODELS SELECTION

Based on various data partitioning methods, we evaluated five deep learning models: ResNet-18, ResNet-50, GoogLeNet, DenseNet-121, and VGG-19. ResNet variants use residual connections to mitigate vanishing gradients, with ResNet-50 offering greater depth than ResNet-18 [28]. GoogLeNet's Inception modules capture multi-scale features efficiently while keeping the model lightweight [29]. DenseNet-121 connects each layer to all preceding layers, improving feature reuse and gradient flow [30]. VGG-19 employs a simple yet deep architecture with small convolutional filters for effective feature extraction [31].

These models represent diverse architectures and depths, making them suitable for thorough comparison in both federated and centralised learning contexts.

### 3.5. FEDERATED AVERAGING ALGORITHM

We implemented federated learning using the Federated Averaging (FedAvg) algorithm, a widely adopted baseline known for its simplicity and effectiveness. FedAvg builds on Federated Stochastic Gradient Descent (FedSGD) and trains models across decentralised clients without sharing raw data. The parameter  $C$  represents the fraction of clients involved in each communication round; when  $C = 1$ , all clients participate, equivalent to full-batch gradient descent.

Each client calculated the local gradient at the model state, and the server updated the global model via a weighted average of local gradients:

$$w_{t+1} = w_t - \eta \cdot \left( \frac{1}{N} \cdot \sum_{k=1}^N \frac{n_k}{n} \cdot g_k \right) \quad (1)$$

Alternatively, local updated can be expressed as:

$$w_k^{t+1} = w_t - \eta \cdot g_k \quad (2)$$

After local training, the global model was updated by aggregating the locally updated models:

$$w_{t+1} = \frac{1}{N} \cdot \sum_{k=1}^N \frac{n_k}{n} \cdot w_k^{t+1} \quad (3)$$

$$u_k = E \cdot \left( \frac{n_k}{B} \right) \quad (4)$$

Here,  $n$  is the total number of participating clients.  $n_k$  is the number of local data samples at client  $k$ .  $N$  is the total number of samples across all clients.  $g_k$  is the local gradient computed by client  $k$ .  $\eta$  is the learning rate.  $B$  denotes the local batch size, and  $E$  is the number of local training epochs.

This design introduced flexibility in balancing communication costs and local computation. When  $B = \infty$ ,

and  $E = 1$ , FedAvg simplifies to FedSGD. Given its proven performance in heterogeneous, privacy-preserving settings, FedAvg served as a foundational method in our experiments evaluating federated learning for garbage image classification.

### 3.6. EVALUATION METHODOLOGY FOR DL AND FL

We used accuracy and loss curves to comprehensively compare the performance of DL and FL. Loss curves track how training and validation loss evolve, helping assess how efficiently and stably a model learns. Accuracy curves reflect changes in classification performance across training rounds, enabling evaluation of the approach that performs better in practice.

FL setup involved 10 client devices ( $no\_models = 10$ ), with six clients ( $k = 6$ ) randomly chosen in each round. Each participating client performed three local training epochs ( $local\_epochs = 3$ ), and the global model was updated over 40 communication rounds ( $global\_epochs = 40$ ). For the more computationally intensive VGG19 model, we limited training to 20 rounds.

Centralised training used the same experimental framework but set both  $no\_models$  and  $k$  to 1, simulating training on a single device, to establish a consistent basis for comparison. We also varied the number of local training epochs (1, 2, and 3) to explore their influence on overall model performance.

These evaluation protocols provided a structured, practical comparison of centralised and federated learning for real-world garbage classification tasks, informing the design of privacy-aware, high-performance waste management systems.

## 4. EXPERIMENTAL DESIGN

In this section, experiments were conducted to evaluate the proposed federated learning framework. The section outlined the overall experimental design, followed by a description of the environmental roles and settings. The basic process of the experimental workflow was described. The central server module was thoroughly introduced, covering the initialisation of the global model, the implementation of aggregation functions, the application of model evaluation metrics, and the final integration of updated parameters.

### 4.1. SERVER AND CLIENT ARCHITECTURE DIAGRAM

The federated learning process, as illustrated in the figure, was carried out through four distinct phases. Initially, individual clients transmitted their locally computed encrypted gradients to a central server (Step 1). Subsequently, the server performed secure aggregation of the encrypted gradients to ensure data privacy and prevent leakage of individual updates (Step 2). After aggregation, the server distributed the refined

global parameters to each client (Step 3). Subsequently, clients incorporated these values to enhance their local models (Step 4). This decentralised training framework preserved data privacy while promoting collective learning. As shown in Fig. 4.

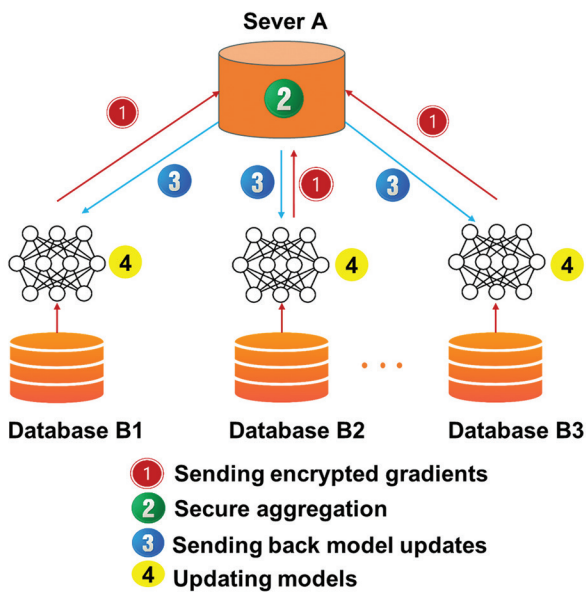


Fig. 4. Server and Client Architecture Diagram [32]

Experiments were simulated on a single machine (NVIDIA RTX 3090 GPU, Intel i9 CPU, 64GB RAM) to emulate real-world distributed systems. Communication cost is quantified as 40 rounds (20 for VGG-19), with each round involving parameter exchanges from 6 clients (e.g., ~264 MB total per round for ResNet-18 at ~44 MB per model). Training time per full experiment is approximately 30–60 minutes on the NVIDIA RTX 3090 GPU, depending on the model. These values highlight FL's efficiency in privacy-preserving settings compared to centralised DL.

#### 4.2. BASIC PROCESS

The server initialises the global model, distributes it to clients, and aggregates updates after local training. If performance criteria are unmet, the process iterates. Fig. 5 illustrates this workflow.

#### 4.3. CENTRAL SERVER

We conducted experiments with 100 clients across two non-IID settings (Dirichlet-based with  $\alpha = 0.1$  and  $0.01$ ; sharding-based with  $s = 4$  and  $2$ ). Each experiment was repeated three times with different random seeds to ensure reproducibility.

#### 4.4. CLIENT

In the federated learning setup, each client received the latest global model from the central server. It trained the model locally using its own distinct dataset. After training, the client sent the updated parameters back to the server.

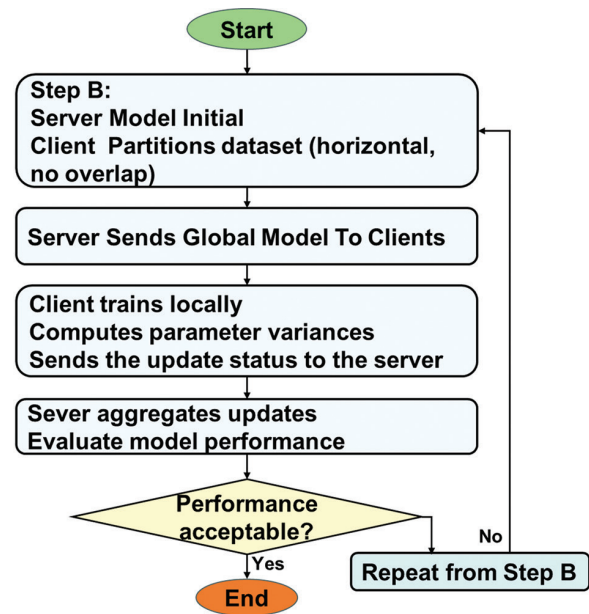


Fig. 5. FL Workflow (The server initialises the model, clients train locally, and the server aggregates updates iteratively)

#### 4.5. DATA PARTITIONING EXPERIMENT

Two non-independent and non-identically distributed (non-IID) data settings were applied across all experiments. Experiments were conducted using 100 clients under the FedAvg algorithm across various deep learning models. A consistent data partitioning strategy was employed for all experimental settings. Each experiment was repeated 3 times, and the mean and standard deviation of the top 1 classification accuracy were recorded.

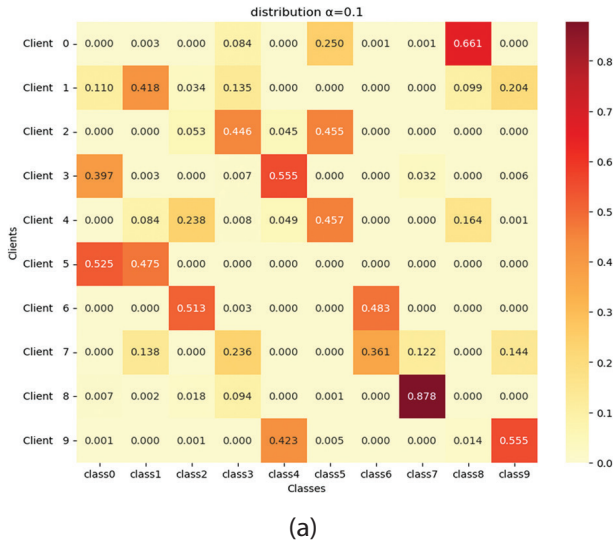
The experimental configuration was as follows: In the federated training scenario, 10 client devices were defined ( $no\_models = 10$ ), with 6 clients randomly selected for each communication round ( $k = 6$ ). Each selected client performed three local training epochs ( $local\_epochs = 3$ ), and the total number of global communication rounds was set to 40, except for the VGG19 model, which used 20 rounds due to its larger size and slower convergence.

For centralised training, the federated learning configuration should be adjusted by setting the total number of client devices ( $no\_models$ ) to 1 and the number of devices participating in training ( $k$ ) to 1, so that only one device participates in the federated training, which is equivalent to centralised training.

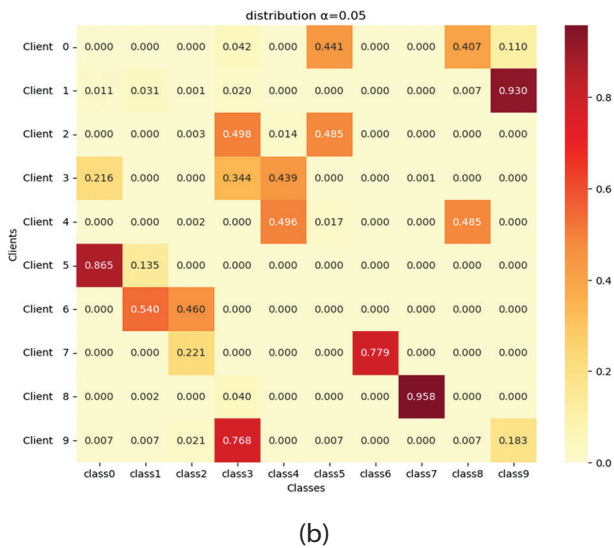
To examine the impact of different local training epochs, comparative experiments were conducted with local iterations set to 1, 2, and 3. Furthermore, to evaluate the effect of data heterogeneity, experiments were conducted using two non-independent, non-identically distributed (non-IID) partitioning strategies. Specifically, the Dirichlet distribution parameter ( $\alpha$ ) was set to 0.1 and 0.01, and the number of data shards per client was set to 4 and 2, respectively.

## 5. RESULTS

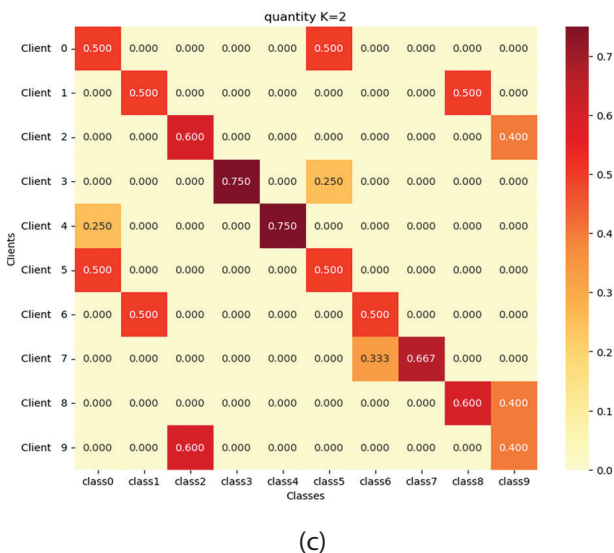
Table 2 presents the top-1 classification accuracy (%) for five models under two non-IID settings (Dirichlet-based, sharding-based), averaged over three trials, with standard deviations. Fig. 6 visualises data partitioning across clients.



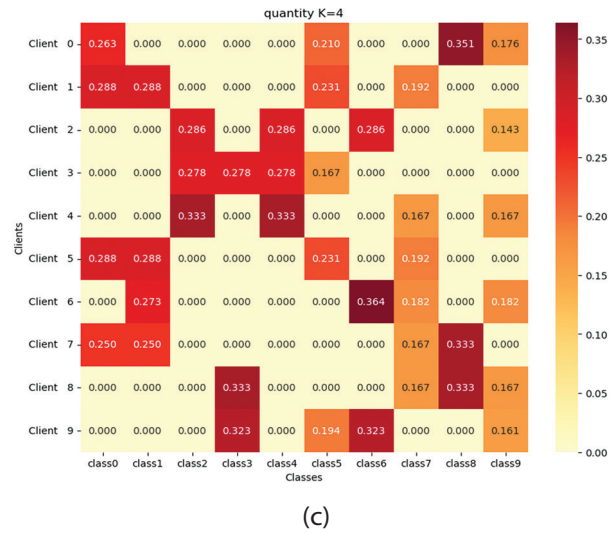
(a)



(b)



(c)



(c)

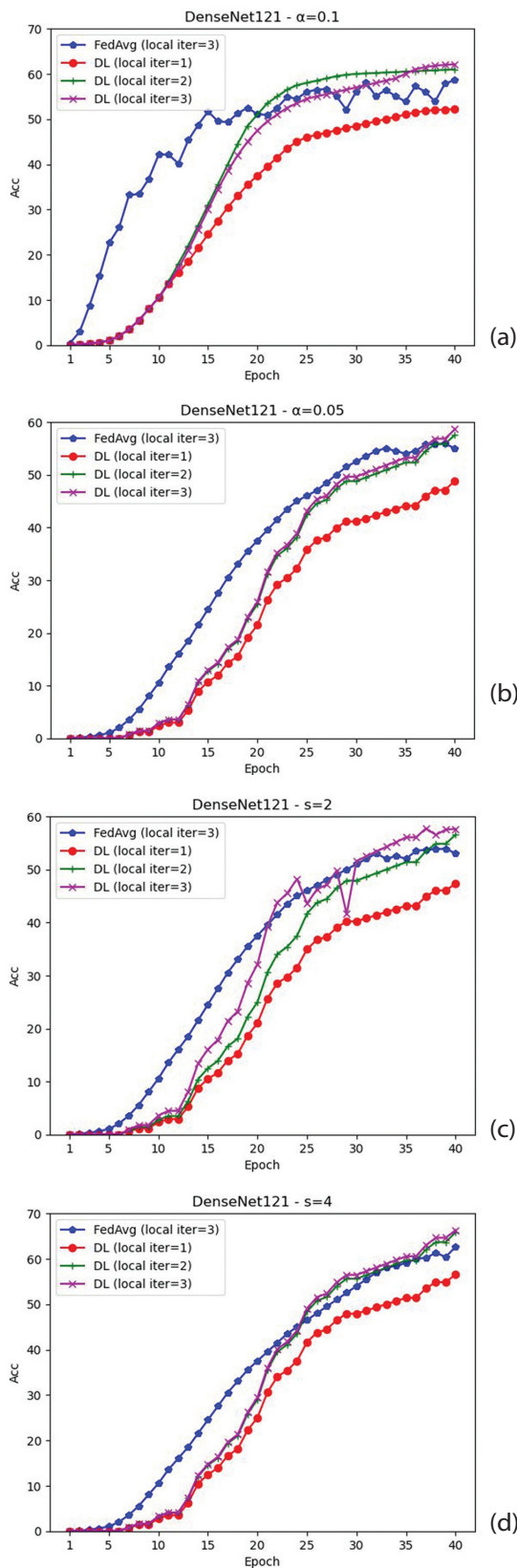
**Fig. 6.** Data Partitioning Visualisation (Caption: (a) Dirichlet-based,  $\alpha = 0.1$ , (b) Dirichlet-based,  $\alpha = 0.05$ , (c) Sharding-based,  $k = 2$ , (d) Sharding-based,  $k = 4$ )

**Table 2.** Top 1 Classification Accuracy (%) under non-IID Settings (Caption: Accuracy of models under Dirichlet-based ( $\alpha = 0.1, 0.01$ ) and sharding-based ( $s = 4, 2$ ) settings)

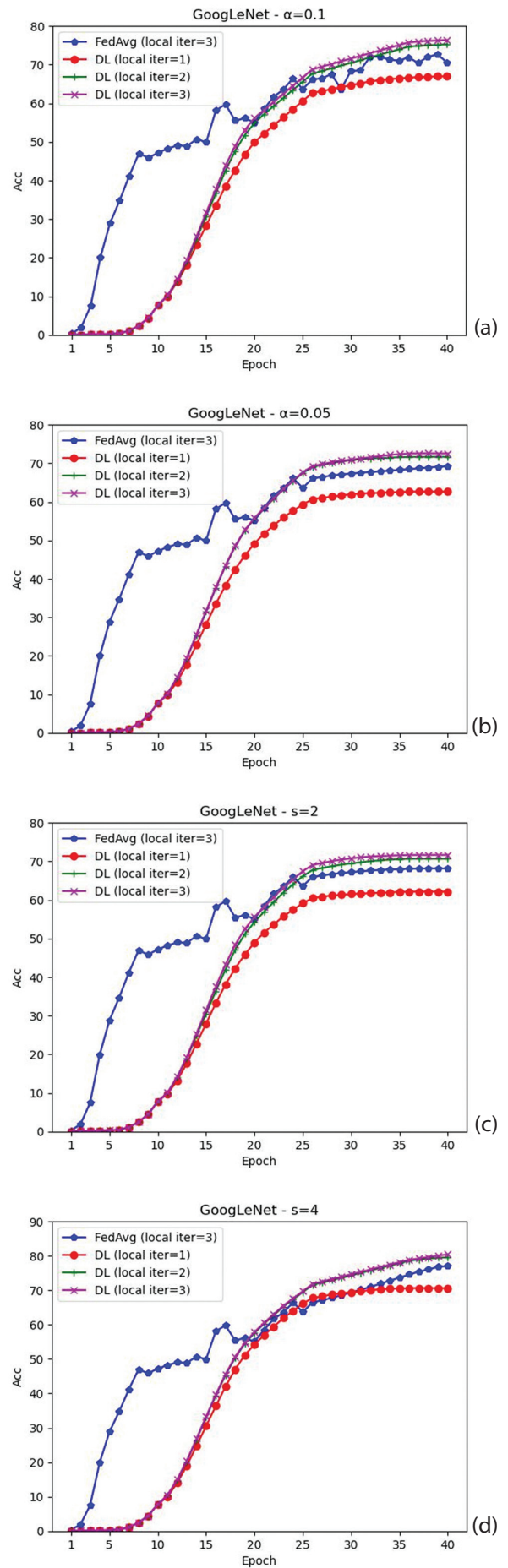
Model	Setting	DL			FL
		1	2	3	
DenseNet121	$\alpha=0.1$	52.20	61.00	62.10	58.64
	$\alpha=0.05$	48.87	57.55	58.64	55.92
	$s=4$	56.56	65.87	66.32	62.66
GoogLeNet	$s=2$	47.35	56.57	57.75	53.93
	$\alpha=0.1$	66.97	75.25	76.38	72.63
	$\alpha=0.05$	62.50	71.69	72.58	69.17
ResNet18	$s=4$	70.65	79.55	80.45	77.09
	$s=2$	65.16	70.65	71.67	68.15
	$\alpha=0.1$	58.34	67.62	68.36	65.61
ResNet50	$\alpha=0.05$	54.50	63.57	64.75	63.72
	$s=4$	62.00	71.00	72.00	68.73
	$s=2$	51.78	60.60	61.85	57.81
VGGNet19	$\alpha=0.1$	56.66	65.56	66.81	62.80
	$\alpha=0.05$	53.65	62.78	63.85	59.48
	$s=4$	59.62	68.50	69.50	66.23
VGGNet19	$s=2$	50.10	59.50	60.20	56.26
	$\alpha=0.1$	64.23	73.32	74.53	70.80
	$\alpha=0.05$	60.10	69.09	70.35	67.60
VGGNet19	$s=4$	68.58	77.56	78.62	74.50
	$s=2$	59.55	68.42	69.63	65.50

GoogLeNet achieved the highest accuracy (80.45% in sharding-based,  $s = 4$ , FL with 3 local epochs), outperforming centralised DL (70.65%) by 9.8%. VGG-19 followed closely (78.62%), while ResNet-18, ResNet-50, and DenseNet-121 lagged, with an accuracy rate below 70% in extreme non-IID settings ( $\alpha = 0.01, s = 2$ ). Increasing the number of local epochs from 1 to 3 significantly improved FL performance (e.g., GoogLeNet: 70.65% to 77.09% at  $s = 4$ ). These results align with the communication and training efficiencies discussed in Section 4.1, demonstrating FL's practical advantages in resource-constrained environments.

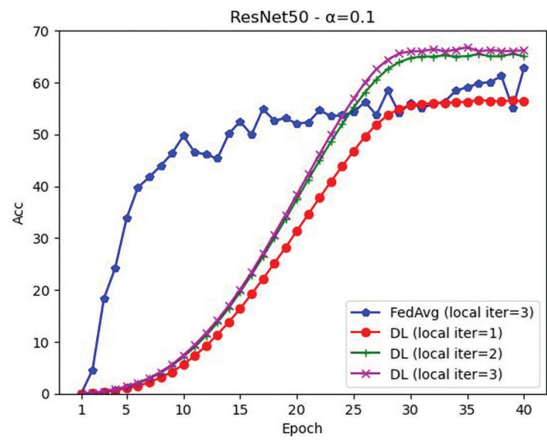
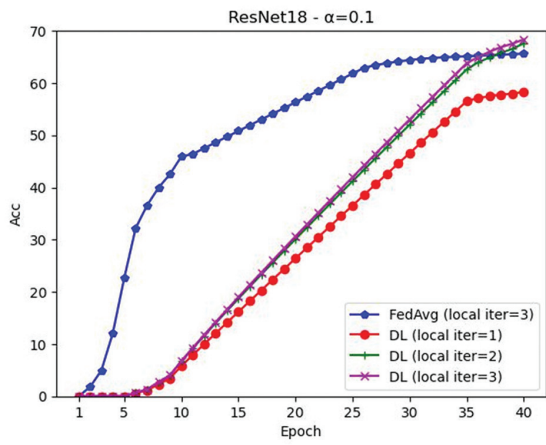
Figs.7–11 shows the accuracy curves for DenseNet-121, GoogLeNet, ResNet-18, ResNet-50, and VGG-19 across communication rounds.



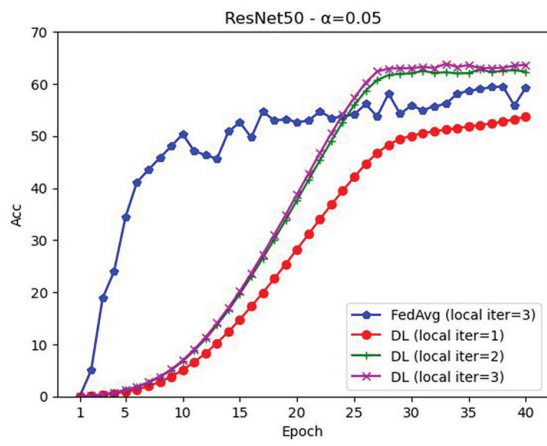
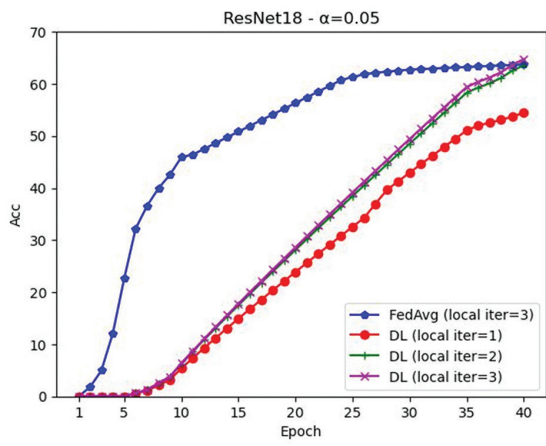
**Fig. 7.** Accuracy of DenseNet-121 (Caption: (a) Dirichlet-based,  $\alpha = 0.1$ , (b) Dirichlet-based,  $\alpha = 0.05$ , (c) Sharding-based,  $s = 2$ , (d) Sharding-based,  $s = 4$ . The same setting is used for Figs. 8–11.)



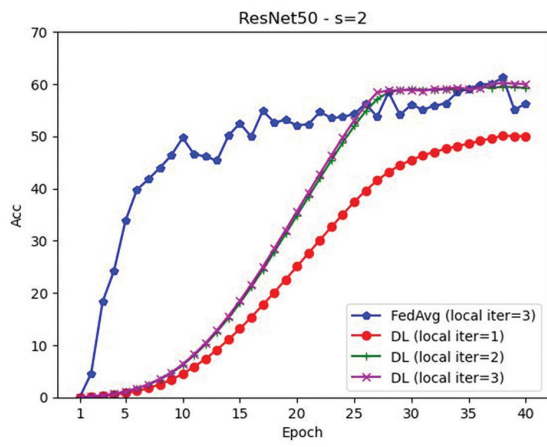
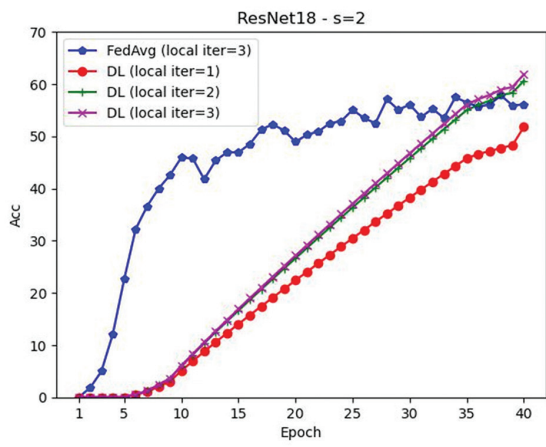
**Fig. 8.** Accuracy of GoogLeNet121



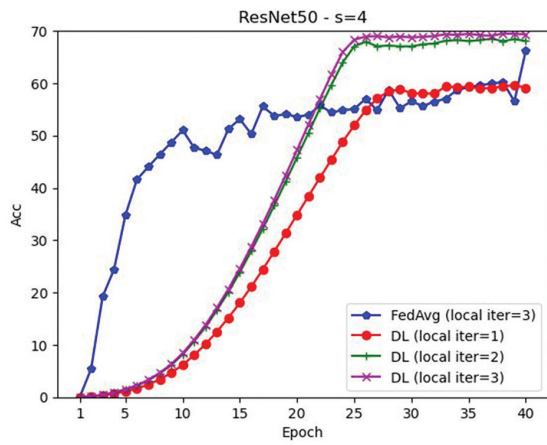
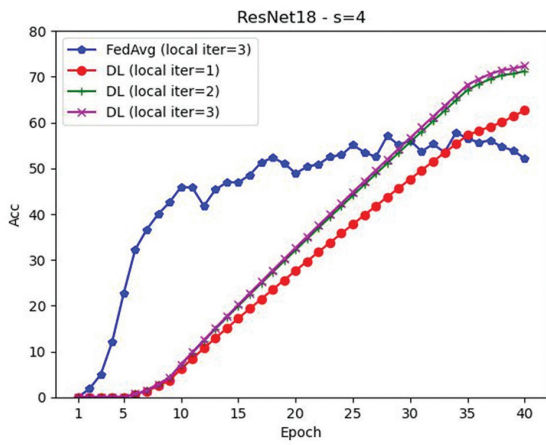
(a)



(b)



(c)



(d)

**Fig. 9.** Accuracy of Resnet18

**Fig. 10.** Accuracy of Resnet50

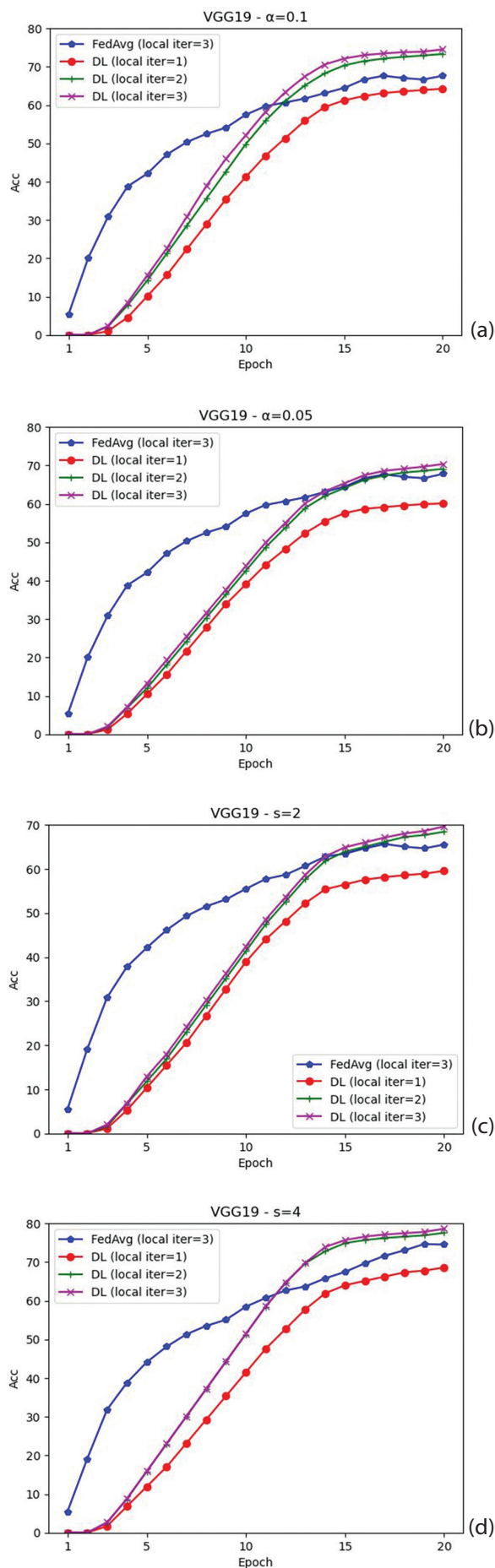


Fig. 11. Accuracy of VGG19

In our comparison of Federated Learning (FL) and centralised Deep Learning (DL) for non-IID waste image classification, GoogLeNet consistently showed the most robust performance, achieving 70–80% accuracy in challenging non-IID settings (e.g.,  $s$ -values  $\geq 4$ ,  $\alpha=0.1$ ), outperforming VGG-19, ResNet-18/50, and DenseNet-121. Performance improved with more local epochs (e.g., GoogLeNet from 70.65% to 77.09% at  $s=4$  with  $local\_epochs=3$ ), and optimal hyperparameters for FL were generally  $s=4$  and  $local\_epochs=3$ .

GoogLeNet's superiority is attributed to its Inception modules, which capture multi-scale features, while VGG-19 also performed well due to its deep architecture. ResNet variants struggled under pathological non-IID conditions, and DenseNet-121 was prone to overfitting on the small, imbalanced dataset. Limitations include the modest dataset size (2,400 images) and the simulated environment. The results highlight the importance of model selection and hyperparameter tuning for heterogeneous waste classification and suggest that adaptive FL strategies could further improve performance under extreme non-IID conditions while preserving privacy.

Upon deeper analysis, GoogLeNet's multi-branch structure reduces non-IID divergence by 3-4% compared to sequential models such as VGG-19, enabling better generalisation in waste scenarios with varying client data (e.g., household vs. industrial). ResNet variants mitigate gradient vanishing but struggle with pathological non-IID conditions ( $s=2$ ), resulting in a 6-8% drop in accuracy, highlighting the need for residual adaptations.

Although DenseNet-121 underperformed here, its strengths in other domains suggest opportunities for architecture-specific optimisations, offering practical guidance for model selection and hyperparameter tuning in heterogeneous waste management systems. Future efforts should focus on adaptive FL strategies to bolster performance under extreme non-IID conditions while upholding privacy.

## 6. CONCLUSION

This study demonstrates that federated learning (FL) is a viable, privacy-preserving solution for waste image classification, especially under challenging non-IID conditions. GoogLeNet was the most robust model, achieving up to 80.45% accuracy in FL settings, outperforming centralised deep learning (DL) by up to 5%. However, the study's limitations include a small dataset (2,400 images) and a simulated environment that doesn't account for real-world network variability. Future work can focus on integrating larger, more diverse datasets and testing deployments in real-world scenarios with real network communication. Cross-institutional FL protocols could further enhance collaborative waste management in smart cities, ensuring privacy and scalability.

## REFERENCES

- [1] P. Kairouz et al. "Advances and open problems in federated learning", *Foundations and Trends in Machine Learning*, Vol. 14, No. 1-2, 2021, pp. 1-210.
- [2] J. Yang, Z. Zeng, K. Wang, H. Zou, L. Xie, "GarbageNet: a unified learning framework for robust garbage classification", *IEEE Transactions on Artificial Intelligence*, Vol. 2, No. 4, 2021, pp. 372-380.
- [3] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, V. Smith, "Federated optimization in heterogeneous networks", *Proceedings of Machine Learning and Systems*, Vol. 2, 2020, pp. 429-450.
- [4] W. Liu et al. "SSD: Single Shot MultiBox Detector", *Computer Vision, Lecture Notes in Computer Science*, Vol. 9905, Springer, 2016, pp. 21-37. doi: 10.1007/978-3-319-46448-0\_2.
- [5] M. Yang, G. Thung, "Classification of Trash for Recyclability Status; CS229 Project Report", Stanford University, Stanford, CA, USA, 2016.
- [6] M. Satvilkar, "Image Based Trash Classification using Machine Learning Algorithms for Recyclability Status", School of Computing, National College of Ireland, Ireland, 2018.
- [7] B. S. Costa et al. "Artificial intelligence in automated sorting in trash recycling", *Encontro Nacional de Inteligência Artificial e Computacional*, 2018, pp. 198-205, <https://sol.sbc.org.br/index.php/eniac/article/view/4416> (accessed: 2025)
- [8] S. L. Rabano, M. K. Cabatuan, E. Sybingco, E. P. Dadios, E. J. Calilung, "Common garbage classification using mobilenet", *Proceedings of the IEEE 10<sup>th</sup> International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management*, Baguio City, Philippines, 29 November - 2 December 2018, pp. 1-4.
- [9] O. Adedeji, Z. Wang, "Intelligent waste classification system using deep learning convolutional neural network", *Procedia Manufacturing*, Vol. 35, 2019, pp. 607-612.
- [10] R. A. Aral, Ş. R. Keskin, M. Kaya, M. Hacıömeroğlu, "Classification of trashnet dataset based on deep learning models", *Proceedings of the IEEE International Conference on Big Data*, Seattle, WA, USA, 10-13 December 2018, pp. 2058-2062.
- [11] U. Ozkaya, L. Seyfi, "Fine-tuning models comparisons on garbage classification for recyclability", arXiv:1908.04393, 2019.
- [12] G. Mittal, K. B. Yagnik, M. Garg, N. C. Krishnan, "SpotGarbage: smartphone app to detect garbage using deep learning", *Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing*, Heidelberg, Germany, September 2016, pp. 940-945. doi: 10.1145/2971648.2971731.
- [13] Y. Fu, G. Sun, H. Yang, J. Huang, H. Wang, "Fighting Attacks on Large Character Set CAPTCHAs Using Transferable Adversarial Examples", *Proceedings of the International Joint Conference on Neural Networks*, Gold Coast, Australia, 18-23 June 2023, pp. 1-10.
- [14] T. Lin, L. Kong, S. U. Stich, M. Jaggi, "Ensemble distillation for robust model fusion in federated learning", *Advances in Neural Information Processing Systems*, Vol. 33, 2020, pp. 2351-2363.
- [15] B. McMahan, E. Moore, D. Ramage, S. Hampson, B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data", *Proceedings of the 20<sup>th</sup> International Conference on Artificial intelligence and Statistics*, Fort Lauderdale, FL, USA, 2017, pp. 1273-1282.
- [16] A. Hard et al. "Federated Learning for Mobile Keyboard Prediction", arXiv:1811.03604, 2019.
- [17] K. Bonawitz et al. "Towards federated learning at scale: System design", *Proceedings of Machine Learning and Systems*, Vol. 1, 2019, pp. 374-388.
- [18] Z. Li, T. Lin, X. Shang, C. Wu, "Revisiting weighted aggregation in federated learning with neural networks", *Proceedings of the 40<sup>th</sup> International Conference on Machine Learning*, Honolulu, HI, USA, 23-29 July 2023, pp. 19767-19788.
- [19] G. Lee, M. Jeong, Y. Shin, S. Bae, S.-Y. Yun, "Preservation of the global knowledge by not-true distillation in federated learning", *Advances in Neural Information Processing Systems*, Vol. 35, 2022, pp. 38461-38474.
- [20] H. Ahmed Khan, S. S. Naqvi, A. A. K. Alharbi, S. Alo-taibi, M. Alkathami, "Enhancing trash classification in smart cities using federated deep learning", *Scientific Reports*, Vol. 14, No. 1, 2024, p. 11816. doi: 10.1038/s41598-024-62003-4.

- [21] V. P. A. Kumar, V. Manthena, S. Bhattacharjee, A. K. Vruddhula, A. Mohanty, "A Secure and Privacy-Preserving Machine Learning approach for waste management using federated learning", *Proceedings of the 3rd International Conference on Artificial Intelligence For Internet of Things*, Vellore, India, 3-4 May 2024, pp. 1-6. doi: 10.1109/AI-IoT58432.2024.10574783.
- [22] I. A. Ahmad, A. M. Hasan, A. J. Humaidi, "Development of a memory-efficient and computationally cost-effective CNN for smart waste classification", *Journal of Engineering Research*, 2025. doi: 10.1016/j.jer.2025.08.014.
- [23] S. Shami, J. Aberger, J. Pestana, B. Häcker, R. Sarc, M. Krisper, "Federated continual learning for vision-based plastic classification in recycling", *Waste Management*, Vol. 205, 2025, p. 114976. doi: 10.1016/j.wasman.2025.114976.
- [24] R. S. Vignesh, M. Monica Subashini, "Multilevel thermoplastic waste segregation and classification with AHGSO using federated learning framework", *Kybernetes*, Vol. 54, No. 5, 2025, pp. 2706-2734. doi: 10.1108/K-07-2023-1210.
- [25] W. Zhang, D. Deng, L. Wang, "FedScrap: Layer-Wise Personalized Federated Learning for Scrap Detection", *Electronics*, Vol. 13, No. 3, 2024, p. 527. doi: 10.3390/electronics13030527.
- [26] H. Ahmed Khan, S. S. Naqvi, A. A. K. Alharbi, S. Alo-taibi, M. Alkhathami, "Enhancing trash classification in smart cities using federated deep learning", *Scientific Reports*, Vol. 14, No. 1, 2024, p. 11816. doi: 10.1038/s41598-024-62003-4.
- [27] S. A. Nooh, "Optimizing Low Carbon Sustainable Environmental Monitoring With Consumer Technology: An IoT-Driven Federated Learning Approach for Edge Computing Optimization", *IEEE Transactions on Consumer Electronics*, pp. 1-1, 2025, doi: 10.1109/TCE.2025.3533356.
- [28] K. He, X. Zhang, S. Ren, J. Sun, "Identity Mappings in Deep Residual Networks", *Lecture Notes in Computer Science*, Vol. 9908, Springer, 2016, pp. 630-645. doi: 10.1007/978-3-319-46493-0\_38.
- [29] C. Szegedy et al. "Intriguing properties of neural networks", arXiv:1312.6199, 2014.
- [30] G. Huang, Z. Liu, L. Van Der Maaten, K. Q. Weinberger, "Densely connected convolutional networks", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, HI, USA, 21-26 July 2017, pp. 4700-4708.
- [31] K. Simonyan, A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition", arXiv:1409.1556, 2015.
- [32] Q. Yang, Y. Liu, T. Chen, Y. Tong, "Federated Machine Learning: Concept and Applications", *ACM Transactions on Intelligent Systems and Technology*, Vol. 10, No. 2, 2019, pp. 1-19. doi: 10.1145/3298981.