CEAS Aeronautical Journal
https://doi.org/10.1007/s13272-025-00865-8

ORIGINAL PAPER

A systematic review of federated statistical heterogeneity in UAV applications

Bara Al-Bataineh[1] · Sirko Schindler[1] · Diana Peters[1] · Marcus Paradies[2] · Matthias Pohl[1]

Received: 10 June 2024 / Revised: 3 June 2025 / Accepted: 5 June 2025

Abstract

This systematic literature review explores Federated Learning (FL) within the context of Unmanned Aerial Vehicle (UAV) applications. FL works by training a global model among clients, where the model is trained locally on each client, and only the model updates are shared. This approach maintains privacy and enables collaborative learning without sharing raw data. The collaborative efforts of multiple UAVs, however, introduce statistical heterogeneity in the collected sensing data due to variations in their respective monitoring areas. In this review, we analyze 31 papers published between 2016 and October 2023. Our review highlights the data properties, FL frameworks, applications, and evaluation methodologies used in these studies. We provide a detailed classification of the current state-of-the-art in FL, particularly focusing on approaches to manage statistical heterogeneity. This review also includes an assessment of the various evaluation methods used in the literature. This review offers a concise overview of the advancements made in addressing statistical heterogeneity in research studies. We will highlight key progress, identify persistent challenges, and explore future research directions. Ultimately, our goal is to provide insights into the ongoing developments in Federated Learning applications for UAV.

Keywords Aviation · Federated learning · Statistical heterogeneity · Non-IID · Machine learning · UAV

1 Introduction

According to Cisco, There will be over 75 billion Internet of Things (IoT) devices in 2025, which is 2.5 times as many as the 31 billion IoT devices that existed in 2020 [1]. IoT devices include sensors, wearable devices, smartphones, connected cars, and Unmanned Aerial Vehicles (UAVs). UAVs, also known as drones, have gained popularity due to their flexibility, line-of-sight (LoS) connections, and 3D mobility [2]. UAVs can be employed in various applications, including military, civil, environmental remote sensing, and agriculture. Also, UAVs are widely utilized for forest fire management, pollution and air quality assessment, coastal ocean observations, cloud and precipitation assessment, and severe storm monitoring applications [3]. Their adaptability and versatility make UAVs a valuable asset in various industries and applications.

The adoption of artificial intelligence, particularly machine learning (ML) techniques, has gained traction in enhancing UAV capabilities [4]. Traditionally, ML techniques heavily depend on cloud processing resources while UAVs assume the role of data collectors [5]. UAVs carry various types of equipment, such as sensors, cameras, and communication devices [6]. They collect data through their sensors and then transmit the data to the cloud for processing and modeling. Furthermore, they can also assist IoT devices that are incapable of transmitting data over a long distance due to energy constraints. UAVs have the ability to dynamically move towards IoT devices, collect the data,

✉ Matthias Pohl
matthias.pohl@dlr.de

Bara Al-Bataineh
bara.al-bataineh@dlr.de

Sirko Schindler
sirko.schindler@dlr.de

Diana Peters
diana.peters@dlr.de

Marcus Paradies
paradies@dbs.ifi.lmu.de

1 Institute of Data Science, German Aerospace Center (DLR), Mälzerstraße 3-5, Jena 07745, Thuringia, Germany

2 Chair of Database Systems and Data Mining, Ludwig Maximilian University Munich, Oettingenstraße 67, Munich 80538, Bavaria, Germany

and transmit it to other devices or the cloud which are out of the communication range of the data-producing IoT devices.

However, there are challenges associated with transmitting data to the cloud. First, the data collected by each UAV potentially cannot be freely shared due to privacy or data protection concerns, since it might contain sensitive information (e.g., security-related observations) [7]. Second, the latency from sending raw data to receiving a decision is unacceptable for some real-time UAV applications (e.g., autonomous drones monitoring and target tracking) [8]. Lastly, the transfer of huge amounts of raw data, such as image and video, to the cloud consumes a lot of bandwidth and energy, which is unacceptable for UAV networks with limited bandwidth and energy supply [9]. Therefore, it would be greatly beneficial if the ML model training could be conducted in a distributive manner in UAV networks directly, without sending data to the cloud. This has led to a growing interest in Federated Learning (FL), which addresses the data privacy concerns and communication overhead associated with centralized model training [10].

Introduced FL was first proposed in 2016 and has been widely used in practice since then [11]. For example, Gboard [12] uses FL to train an ML model to suggest search queries based on the typing context. FedHealth [13], a framework for wearable healthcare, can achieve accurate and personalized healthcare without compromising privacy and security. One of the main objectives of this systematic review is to explore the utilization of FL in UAV applications. The major contributions are as follows: Our major contributions include an extensive review of studies published from 2016 to October 2023 addressing the challenge of non-Independently and Identically Distributed (non-IID) data distributions in FL (Section 2) in the context of UAV applications. The methodology for selecting and analyzing these studies is based on the PRISMA flow diagram [14]. We provide an in-depth summary of the selected studies related to statistical heterogeneity, encompassing aspects, such as the FL architecture, learning processes, environmental attributes, non-IID data distributions, data realism, applications, and their evaluation. Based on the included studies, we identify key open problems and outline potential future research directions.

This review follows a structured organization. Section 2 provides necessary background information while Sect. 3 describes the research method used to conduct the study. In Sect. 4, the focus is on the distribution of training data. Section 5 delves into state-of-the-art frameworks, machine learning models used, and the types of UAV client communication in the aggregation step. Section 6 explores various use cases, and Sect. 7 outlines the evaluation matrices employed. Section 8 outlines the open challenges and future directions encountered throughout the study. Finally, we present the conclusion in Sect. 9.

2 Federated learning

Alsamhi et al. proposed a first server-based FL framework designed for UAV networks (see Fig. 1) which relies on a centralized server that coordinates the FL process [8]. According to them, the process can be summarized as follows:

- ***Step1: Data collection.*** The UAV clients acquire private data through their sensors from various areas.
- ***Step2: Global model broadcasting.*** The UAV leader sends the global model to all UAV clients.
- ***Step3: Local model training.*** UAV clients receive a copy of the global model and employ it along with their local data for training their respective local models.
- ***Step4: Upload local model weights.*** The UAV leader receives the model weights from the participating UAV clients.
- ***Step5: Aggregation.*** The UAV leader aggregates received model weights into a new global model.

The learning process of FL involves minimizing a loss function on each UAV client through a weighted aggregation method (e.g., federated average (FedAvg) [11]). The goal is to minimize the overall objective function [15]:

$$min f(w) = \sum_{k=1}^{K} \frac{n_k}{n} \cdot F_k(w)$$

where

$$F_k(w) = \frac{1}{n_k} \sum_{i \in \mathcal{P}_k} f_i(w),$$

where $f(w)$ denotes the loss function for the global model and w denotes the model parameters. $F_k(w)$ is the local loss function for the k-th UAV client, defined as the average loss over its local dataset. $f_i(w)$ is the loss function for the i-th data sample of the kth UAV client. Let n be the total number of data samples across the UAV clients, where the kth UAV client has a dataset, denoted as P_k, consisting of n_k data samples.

In each iteration of the FL process, each UAV client aims to minimize its local loss function using its local data.

2.1 Non-IID data

When the local data contains non-Independently and Identically Distributed (non-IID) attributes, i.e., the distribution of data across UAV clients is not uniform, significant disparities can arise between the local models and the global model [16].

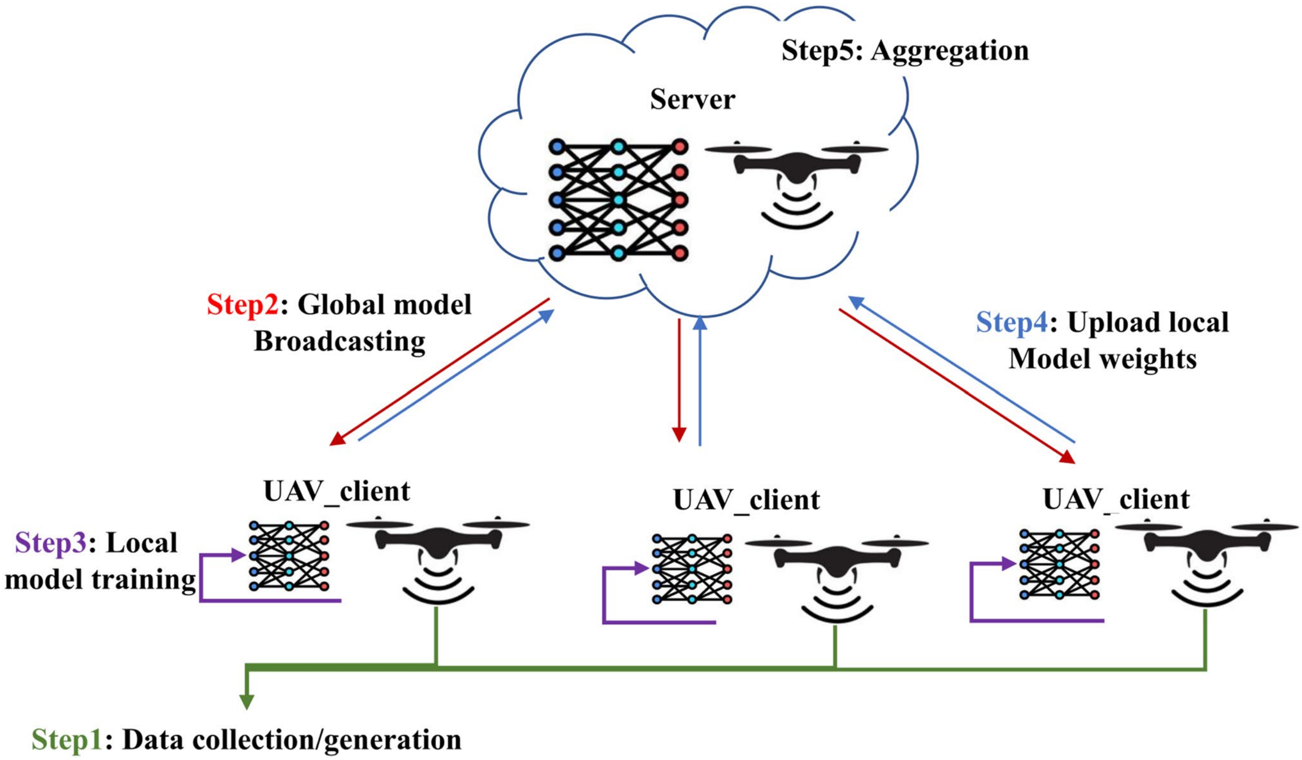

Fig. 1 Federated Learning for UAV computing collaboration

As shown in Fig. 2, in the IID data scenario, the average model w_{t+1}, obtained by aggregating local models, aligns closely with the global optimal model w^*. This alignment occurs because it is equidistant from each local optimal model solution w_1^* and w_2^*. However, with non-IID data, the global optimal solution w^* may be more closely aligned with one of the local optimal models, specifically w_2^*. This results in a discrepancy between the averaged model w_{t+1} and the global optimal solution w^*, as data heterogeneity leads to an uneven influence on the averaged model. If these local model weights are subsequently uploaded to the UAV leader for aggregation, it can negatively impact the accuracy of the global model [16]. UAV clients frequently generate and collect data in a highly non-IID manner across the network. When multiple UAV clients collaborate, the difference in the monitoring area of each UAV client causes statistical heterogeneity in the collected sensing data. There are several ways in which the data among UAV clients can deviate:

- **Label distribution**. The distribution of available labels across different UAV clients is not uniform or balanced. Each UAV client captures specific types of data or objects, resulting in variations in labeling schemes among UAV clients. Two different kinds of label distribution have been investigated in the literature: (1) Each UAV client holds data samples with a fixed number of labels, or (2) a portion of the data samples of labels is distributed to UAV clients with a certain probability.
- **Feature distribution**. Features of the collected data differ across UAV clients. For instance, images of the same object may exhibit variations in terms of brightness, occlusion, camera sensor readings, and more.
- **Quantity distribution**. The size of the local dataset varies across UAV clients.

A non-IID data distribution in UAV clients affects the performance of the ML model (i.e., learning accuracy, stability of the FL algorithm, convergence behavior, and communication efficiency) [17]. The objective of this systematic review is to provide a comprehensive and up-to-date exploration of statistical heterogeneity in FL, particularly in the context of UAV applications.

3 Research method

We followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) approach to conduct our systematic review [14]. PRISMA provides a comprehensive set of guidelines to ensure the transparent and rigorous reporting of systematic review processes and findings. By following PRISMA, we maintain a structured

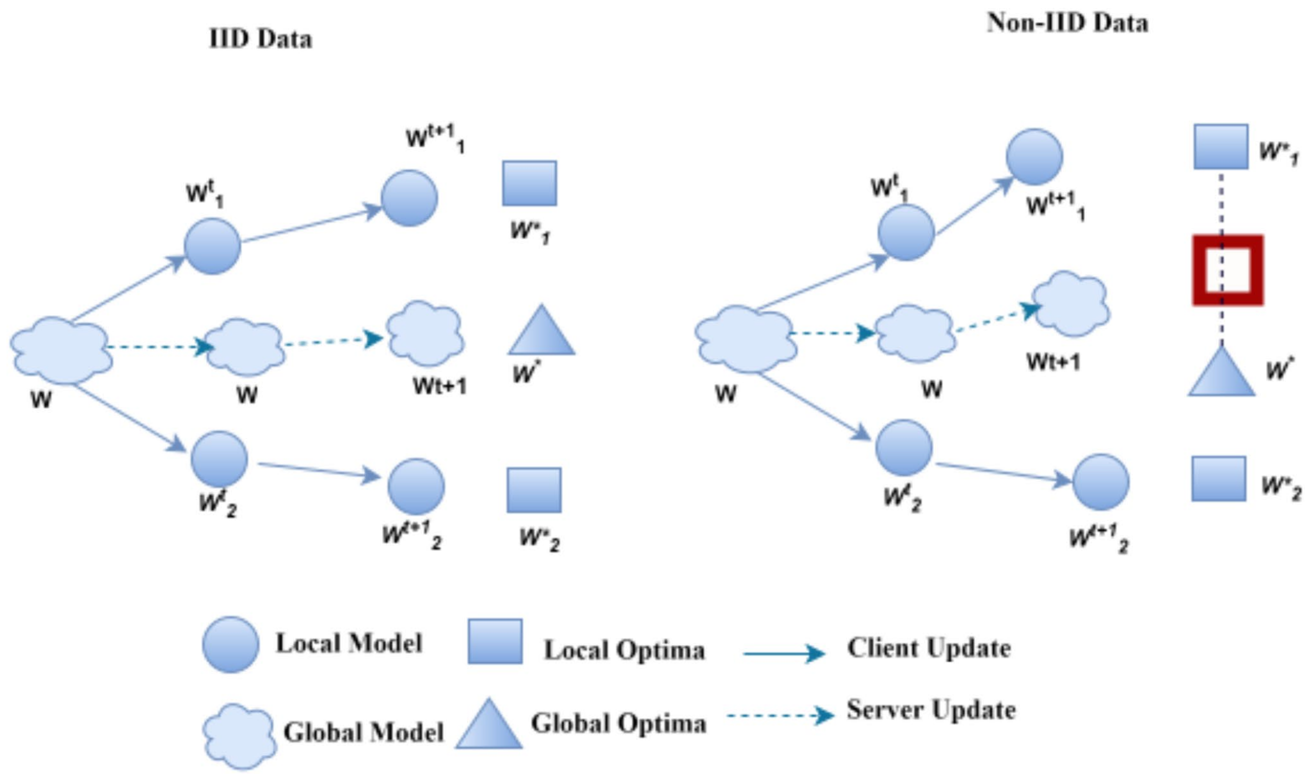

Fig. 2 Effects of IID and non-IID data on FL

and methodical approach throughout our research, from the formulation of research questions and search process to data extraction, quality assessment, and the synthesis of results.

3.1 Research questions

We divide the main objective into the following research questions:

RQ 1 What are specific data characteristics in UAV applications, in particular with respect to non-IID distributions and training data? (Sect. 4)

RQ 2 What are the state-of-the-art FL frameworks applied in UAVs to mitigate the challenges posed by non-IID data distributions and what are their specific capabilities and limitations? (Sect. 5)

RQ 3 In which UAV application scenarios has FL been employed to handle non-IID data distributions? (Sect. 6)

RQ 4 Which evaluation metrics are essential for assessing the performance and efficiency of FL in the presence of non-IID data distributions in UAV applications? (Sect. 7)

RQ 5 What are the open challenges and future directions of FL related to UAV applications? (Sect. 8)

3.2 Search process

In our systematic review, we followed the PRISMA flow diagram [18], shown in Fig. 3, which outlines the review's main steps:

1. In the identification step, we conducted a search covering 1 January 2016 and 2 October 2023, using the following search engines and databases: IEEE Xplore,[1] PubMed,[2] Web of Science,[3] ACM Digital Library,[4] Arxiv,[5] SpringerLink,[6] Scopus,[7] ScienceDirect,[8] and

[1] https://ieeexplore.ieee.org/Xplore/home.js

[2] https://pubmed.ncbi.nlm.nih.gov/

[3] https://www.webofscience.com/wos/woscc/basic-search

[4] https://dl.acm.org/

[5] https://arxiv.org/

[6] https://link.springer.com/

[7] https://www.scopus.com/search/form.uri?display=basic#basic

[8] https://www.sciencedirect.com/

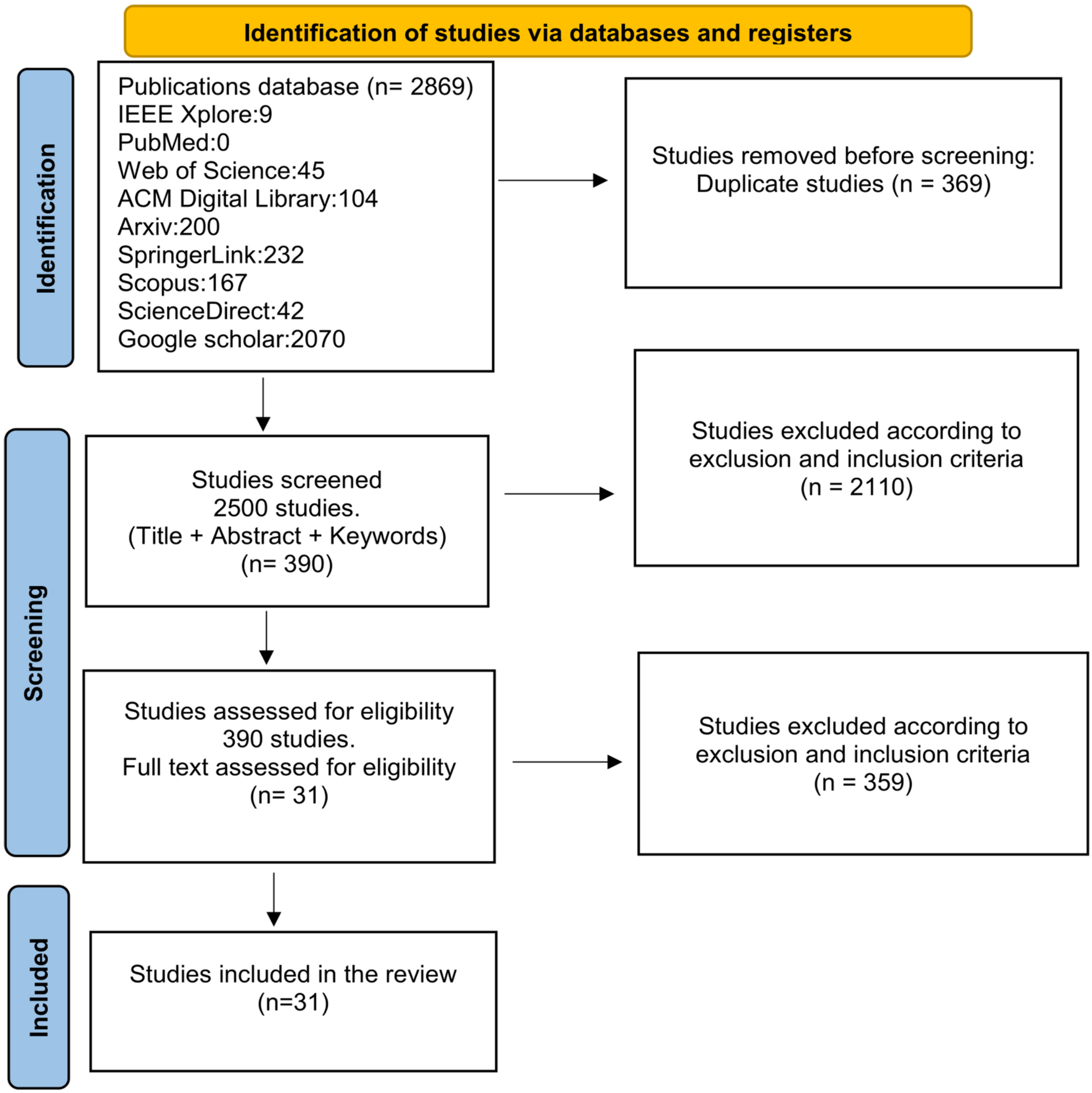

Fig. 3 Study selection using PRISMA flow diagram method consisting of identification step, screening step, and included step

Google scholar.[9] Our search strategy revolved around the terms “federated learning”, “statistical heterogeneity”, and “Unmanned Aerial vehicle” and included synonyms and abbreviations as supplementary terms consisting of "federated ML", "federated artificial intelligence", "federated AI", "non-IID", "UAV", and "Edge Device" to increase the search results. This initial search resulted in 2869 studies satisfying the search criteria. Then, 369 studies were removed due to duplications, ending with 2500 result studies in the identification step.

2. In the screening step, we reviewed 2500 resulting studies based on their titles, abstracts, and keywords, using the following inclusion and exclusion criteria. We excluded 2110 studies based on the following exclusion criteria: (i) non-relevance of the focused subject, (ii) categorized as books, dissertations, or theses, (iii) identified as survey articles, or (iv) written in a non-English language.

[9] https://scholar.google.com/

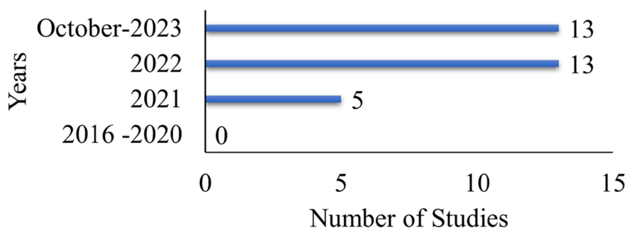

Fig. 4 Distribution of selected studies by publication year

We ended up with 390 studies that (i) discuss the statistical heterogeneity in FL for UAV applications, and (ii) are published in English for further full-text assessment.

Next, 359 studies were excluded based on the exclusion criteria during the full-text assessment.

3. In the inclusion step, **31 studies** summarized in Table 4 (see Appendix) were selected for further analysis, and their results are discussed in this review. Table 5 (see Appendix) presents the list of abbreviations and synonyms used in this systematic review.

3.3 Analysis and synthesis

An initial analysis of the included studies shows that statistical heterogeneity in FL for UAV applications is a rather recent research area. Figure 4 illustrates this aspect with the distribution of included studies by publication year. We observe that from the studies selected 15% were published in 2021, 44% were published in 2022, and 44% of studies published in 2023 before October. This shows that the publication rate and field of interest have grown significantly. We classified the included studies into the following topics within the context of statistical heterogeneity in FL for UAV applications:

1. Data properties (see Sect. 4)
    - Non-IID Distributions
    - Training Data
2. FL Frameworks (see Sect. 5)
    - State-of-the-Art Methods
    - Client–Server Communication
3. Applications (see Sect. 6)
    - Use cases
4. Evaluation Metrics (see Sect. 7)
    - Accuracy
    - Convergence Analysis

Table 1 Summary of non-IID distribution categories employed in the FL for UAV applications

Non-IID distribution categories	Studies
Label distribution	[5, 19–32]
Quantity distribution	[29, 33–37]
Feature distribution	[5, 33, 34]
Mixed-type (Label and Quantity) Distribution	[19, 21, 38–40]
Mixed-type (Label and Feature) distribution	[41]
Undefined non-IID distribution	[2, 42–47]

The next section explores the studies that focus on each of these topics considering the research questions proposed for this review.

4 Data properties

This section aims to include an in-depth analysis of various non-IID data distribution utilizing FL techniques in UAV applications.

4.1 Non-IID distributions

The training data on each UAV client heavily depends on the usage of particular local devices, and therefore, the non-independently and Identically Distributed (non-IID) data distributions of UAV clients may be entirely different from each other. For example, each UAV client collects data from different regions, experiences varying weather conditions, and employs different sensors or data collection methodologies.

Based on the examined studies, we derive the following four categories of non-IID data: (1) quantity distribution, (2) label distribution, (3) feature distribution, and (4) mixed-type distribution and summarize the categorization in Table 1.

Almost half of all studies have focused on label distribution among UAV clients. Label distribution has been determined using one of two methods: First, Dirichlet distributions which are commonly used as prior distribution in Bayesian statistics [48] and are chosen to simulate real-world data distributions. Here, one can adjust the imbalance level by varying the alpha parameter where smaller values lead to more unbalance distributions [19, 21, 24, 35]. Second, the process involves sorting labeled data and dividing it into shards among UAV clients [21, 22]. One of the studies focused on quantity distribution and determined it through random sampling from a Gaussian distribution [34]. The remaining studies, which did not specify the non-IID distribution category, used random partitioning instead.

4.2 Training data

The impact of a non-IID data distribution on the performance of the global model is of critical concern, as highlighted by the studies. For example, a high sensor noise level in a local device can significantly degrade the quality of the local dataset [33]. This degradation subsequently results in a decline in the precision of the global model, which ultimately affects the quality of model predictions. Moreover, non-IID data can slow down convergence rates in learning models, making the training process less efficient and more time-consuming. This, in turn, impacts the accuracy of the global model, which becomes skewed [49]. These observations emphasize the paramount importance to address non-IID data distribution challenges to ensure optimal model performance and robustness.

Mitigating the impact of non-IID data distributions in UAV applications is vital for enhancing the global model performance. Thus, the important step is to represent real-world UAV scenarios accurately using realistic datasets. By selecting or creating datasets that closely mimic the data collected by UAVs during their missions, we can improve the quality of training data. This can involve incorporating diverse environmental conditions, various flight scenarios, and factors like sensor noise and different data collection methodologies into the dataset. By enhancing the realism of datasets, we can work towards minimizing the negative effects of non-IID data distribution and ensuring the success of UAV applications in real-world scenarios.

Table 2 provides an overview of the datasets used in studies, along with corresponding paper references, and classifies these datasets based on realism. Additionally, the table displays the ML models used for training these datasets.

4.3 Characteristics of realistic datasets

Unfortunately, a generic framework to characterize the realism of a dataset could not be identified in scientific research. However, the prevalence of challenges stemming from unrealistic data significantly undermines the integrity of the research items collected.

Many datasets are designed specifically for algorithm benchmarking rather than reflecting actual data collection processes. In the context of UAV networks, simulated environments may not accurately capture the complexities of real-world UAV sensor data [22, 33]. Additionally, synthetic training data often overlooks UAV-specific challenges, such as motion blur, vibration effects, and sensor limitations [6]. Standard image datasets like MNIST [62] contain clean, centered, and size-normalized graphics that rarely reflect real-world conditions. In UAV contexts, this issue is magnified as aerial imagery involves complex variations in altitude, angle, lighting, and environmental conditions [6, 54, 55]. However, the data is primarily generated using **simplified distributions**. Traditional benchmark datasets lack natural variations in lighting, background, positioning, and noise. This presents a particular challenge for UAV applications where environmental factors significantly influence data quality. Aerial imagery for scene classification [57] must account for variable altitudes and perspectives. Power line inspection datasets [44, 46] need to capture seasonal variations and weather conditions. Agricultural monitoring [36, 53] requires adaptation to different growth stages and lighting conditions. This clearly highlights a **limited variability** in the data. Non-realistic datasets often focus on narrowly defined problems that do not generalize well. UAV applications frequently operate in heterogeneous environments. For instance, a UAV trained in one geographic area may perform poorly in other regions [27]. Additionally, models developed for specific detection tasks, such as identifying prohibited items, may not be effective when applied in different security contexts [45, 56] (**Domain Specificity**).

Realistic datasets accurately represent the true statistical properties of the phenomena being studied (**Natural Distribution**). In UAV applications, disaster monitoring datasets must encompass genuine disaster scenarios, capturing their full complexity [42, 54]. Natural scene classification requires accurately representing the true distribution of environmental features [55]. Traffic monitoring must reflect actual traffic patterns and diverse weather conditions . Real-world data are often affected by natural noise, class imbalances, and the presence of outliers (**Inherent Complexity**). UAV application datasets must address non-IID data across different UAVs in a network [17, 35, 54], and natural sensor noise from UAV movement and environmental interference [4]. Capturing data as it would be encountered during actual deployment is essential (**Ecological Validity**). UAV federated learning systems encounter unique challenges. Intermittent connectivity issues can disrupt the synchronization of models, leading to inconsistencies that compromise overall performance [28]. Furthermore, energy constraints impose limitations on both computational power and communication capabilities, restricting the ability to process and share data effectively [32, 36, 37]. Additionally, dynamic positioning requirements can hinder the consistency of data collection, making it challenging to maintain reliable and accurate information over time [47]. Accounting for concept drift and changing patterns over time is essential (**Temporal Relevance**). UAV systems should address seasonal variations (e.g., in agricultural monitoring [43, 53]), evolving security threats (e.g., in intrusion detection [32, 51]), and changing environmental conditions (e.g., in disaster response [42, 54]).

The Table 2 indicates that most studies rely on overly simplistic datasets, predominantly the MNIST dataset. We argue that MNIST is unrealistic for unmanned aerial vehicle

Table 2 Overview of Datasets with corresponding ML models and realism

Datasets	References	Realistic	ML model
Cityscapes [50]	[33]	✓	DDRNet [33]
			FCNN [33]
CIC-IDS2017 [51]	[34]	✓	RNN [34]
			RF [34]
			DT [34]
			SVM [34]
			GAN-LSTM [34]
			CGAN-LSTM [34]
			FL-CGAN-LSTM [34]
AIDER [52]	[42]	✓	VGG16 [42]
			ResNet152 [42]
			Inception ResNet [42]
IDC [53]	[43]	✓	Not-mention
FLAME [54]	[36]	✓	CNN [36]
			Xception [36]
Aerial scene classification [55]	[29]	✓	CNN [29]
			ResNet-18 [29]
Dataset generated by the authors from NWPU-RESISC4Google Earth, different public repositories	[41]	✓	R-CNN [41]
			YOLOV3 [41]
PIDray [56]	[45]	✓	YOLOv7 [45]
			SSD [45]
			R-CNN [45]
Dataset generated by the authors using four different UAV models' stock transmissions	[26]	✓	CNN [26]
RSSCN7 [57]	[30]	✓	CNN with BN [30]
			CNN [30]
			ResNet 18 [30]
SVHN [58]	[47]		LeNet [47]
Cifar-10 [59]	[20, 23, 24, 38, 46]		CNN [20, 46]
			RL [38]
			ResNet-9 [24]
Cifar-100 [59]	[23, 35]		CNN [35]
Fashion-Mnist [60]	[5, 24, 25, 27, 35, 47]		CNN [5, 27, 35]
			FCNN [25]
			LeNet [47]
			ResNet-9 [24]
Shakespeare [61]	[5]		LSTM [5]
MNIST [62]	[5, 19–23, 25, 27, 28, 33, 35, 38, 39, 44, 46]		CNN [5, 19, 20, 22, 27, 32, 35, 39, 44, 46]
			FCNN [22, 25, 33]
			RL [38]
			AlexNet [21]
			ResNet-9 [24]
Sent140 [61]	[5]		LSTM [5]

(UAV) applications due to Data Domain Mismatch and inadequate Data Size and Complexity. The MNIST dataset, consisting of small grayscale images of digits, differs greatly from the high-resolution aerial imagery, sensor data, and video streams typical in UAV operations. This domain mismatch hinders the effective training of models on UAV data.

Additionally, with only 60,000 training images and 10,000 test images, the MNIST dataset is too limited for the diverse challenges faced in real-world UAV applications. Smaller datasets like MNIST fail to provide the necessary generalization for these complex tasks, rendering them insufficient for capturing the intricacies of actual UAV data.

4.3.1 Indication of realistic datasets

The creation of a robust assessment framework to evaluate dataset realism remains an important open research question and will be part of future work. However, our research has identified several key indicators that can help gauge dataset realism.

It is essential to assess whether data was collected from genuine UAV operations or from artificially created scenarios. The **data collection process** must effectively capture UAV movement patterns, altitude variations, and sensor limitations, as these factors greatly impact performance and reliability. Further, the dataset must reflect the **distribution** of data in real UAV deployments, capturing the non-IID nature of distributed UAV networks. Various **complexity measures** can effectively characterize the realism of UAV datasets. Assessing the heterogeneity among different devices is crucial, as it impacts performance and interoperability. Metrics such as entropy and class separability can quantify the intrinsic difficulty. Entropy reveals data uncertainty, while class separability measures how well different data categories can be distinguished. Another indicator is the **transferability** of a model performance to real UAV applications while a model is trained on a dataset. Further performance can be identified through cross-domain applicability between different operational environments [24, 27, 30].

5 Federated learning frameworks

Over time, researchers have increasingly directed their focus towards addressing the non-IID data challenge in FL. In this section, we discuss these methods in detail.

5.1 State-of-the-art methods

Over time, researchers have increasingly directed their focus towards addressing the non-IID data challenge in FL. In this section, we discuss these methods in detail.

5.1.1 Data sharing

Data sharing is the method that involves creating a small subset of data that is shared globally among all edge devices (UAV clients and UAV leader). This subset has a uniform distribution over classes, which helps to mitigate non-IID issues. Tursunboev et al. employ commonly shared data constructed offline at the UAV leader by collecting representative data samples from the UAV clients to effectively solve the divergence issue. Additionally, they introduce a hierarchical aggregation of local models from both UAV clients and UAV leader to update the global model [5]. Similarly, Reus-Muns and Chowdhury employ a centralized subset of globally shared data among all UAV clients. They combine this data-sharing method with a weighted loss function that considers the dataset's non-uniform class distribution, thereby scaling the loss for each training unit based on class distribution [26]. The experiment results indicate that already sharing less than 5% of the global data can lead to performance improvements.

5.1.2 Split learning

Split Learning (SL) [9, 63] is a promising variant of FL where the ML model is split into several sub-models with the specific layer known as the cut layer and distributing them to different edge devices which facilitates distributed learning via sharing the cut layer's activations, called smashed data. Liu et al. [21, 22] have developed the algorithm outlined in Fig. 5 to tackle the challenges posed by non-IID data among UAV clients in a wireless network. Their methodology involves selecting a subset of UAV clients to implement SL $(u_n, u_n + 1)$ with lower computational capability. The remaining UAV clients are assigned to FL (u_1, u_N), which involves less communication overhead when the dataset is large.

The learning process involves UAV clients conducting parallel, local model training. For FL, UAV clients (u_1, u_N) receive global model parameters from the Base Station (BS) (UAV leader) (w_t) and proceed with local training on their respective local datasets $D_1, ..., D_N$. In contrast, SL UAV clients $(u_n, u_n + 1)$ receive a sub-model up to the cut layer of the global model from the BS. The outputs of the forward propagation at the cut layer $(a_t^n, a_t^n + 1)$ are sent back to the BS, which completes the remaining forward propagation from the cut layer to the last layer w_t^e and computes gradients $w_t^{n,e}, w_t^{n+1,e}$ from the last layer to the cut layer. These gradients at the cut layer g_t^n, g_t^{n+1}, and only these gradients, are sent back to the UAV client, which then completes the remainder of the back propagation. The aggregation of model updates takes into account both the average local model updates from FL UAV clients $\triangle w_t^1, \triangle w_t^n$ and the updates from SL UAV clients $\triangle w_t^{n,1}, \triangle w_t^{n+1,1}$. The learning process iterates until a desired convergence performance is achieved or the final iteration is reached.

Sun et al. exclusively applied the SL methodology to all UAV clients without adopting a hybrid approach as used

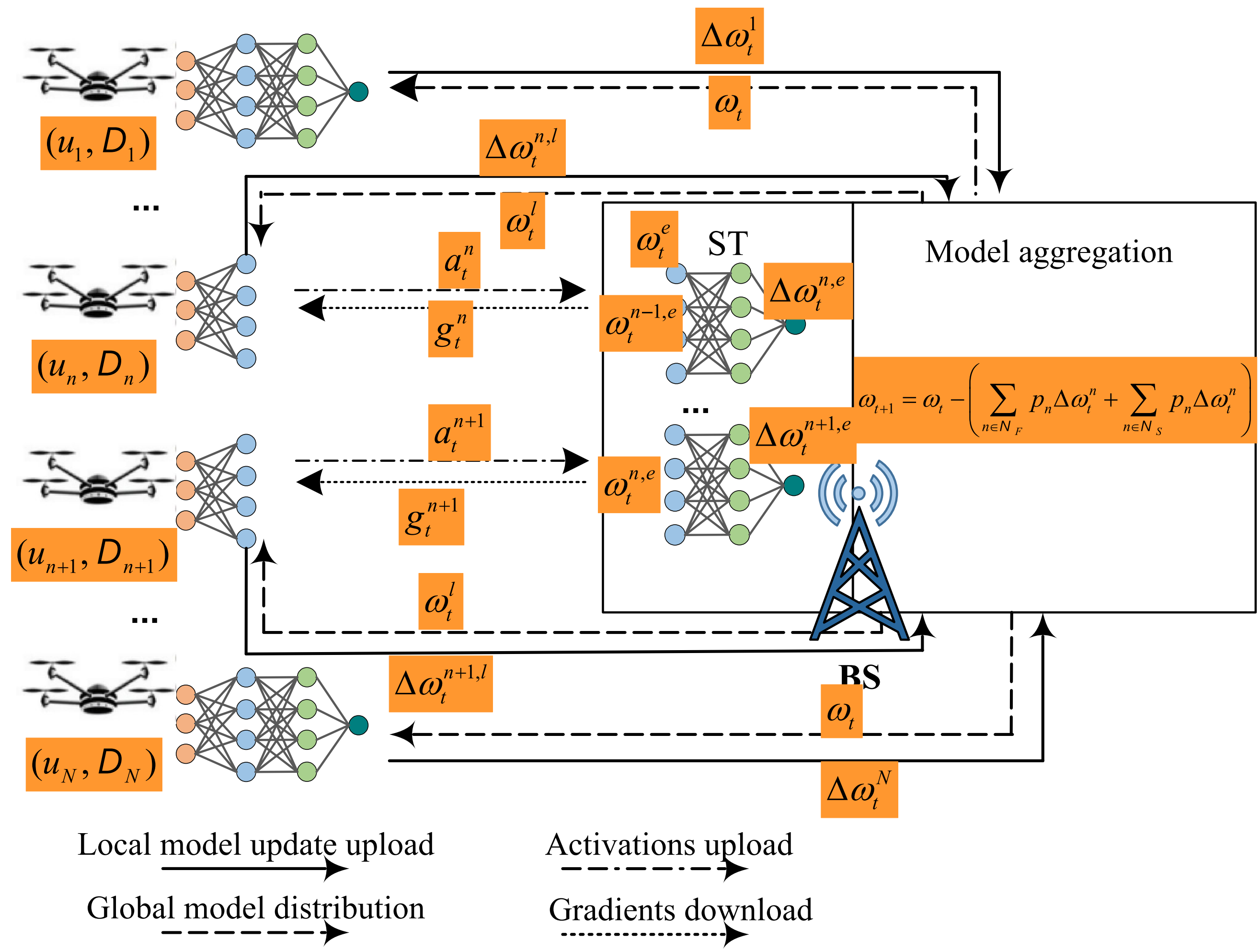

Fig. 5 The learning procedure of the wireless HSFL algorithm [21]

by earlier works. Their focus was on using SL for assisted image classification tasks [29].

5.1.3 Clustering

Hoang et al. presented a novel clustered and decentralized FL framework tailored for UAV swarms. The research [20] introduced an iterative clustering algorithm based on the K-means algorithm to efficiently partition the UAV network into clusters, ensuring connectivity among Cluster-Head (CH) UAVs (UAV leader). Two inter-cluster aggregation schemes, Fully Coordinated Aggregation (FCA) and k-Hop Aggregation (kHA), are proposed and evaluated for different learning scenarios. The evaluation for FCA closely aligns with conventional FL and achieves good performance results, while 1HA (K=1) exhibits delayed convergence, particularly in non-IID distributions.

Pei et al. proposed a novel approach, named Clustered Federated Learning Multi-Classifier, to address the challenges posed by heterogeneous decentralized data in UAV edge devices. The study [23] introduces multiple classifiers to satisfy diverse UAV client needs. To manage the additional storage, computation, and transmission overheads associated with multiple classifiers, the study presents two optimization strategies: first, a Multi-Head Classifier (MH) for sharing feature representations among classifiers, and only multiple branches are introduced at the final classifier layer. Second, the Stochastic classifier (SC) assumes that the classifiers obey a certain prior distribution and then samples the classifiers from it. The evaluation emphasizes the effectiveness of both optimization schemes in mitigating data heterogeneity challenges.

Lin et al. designed a consensus mechanism to mitigate model divergence in Device to Device (D2D) communication within a cluster. Devices (UAV clients) can systematically share their model parameters with others in their neighborhood to form a distributed consensus among each cluster of edge devices. Only one device from the cluster needs to upload the cluster parameter model to the server (UAV leader) during global aggregation, as opposed to the conventional FL architecture where most of the devices are required to upload their local models. They effectively improve performance in the presence of statistical heterogeneity [25].

Wang et al. proposed a comprehensive framework for efficient ML model training using UAV swarms in the context of geo-distributed device clusters. Three types of UAVs are considered within a swarm: leaders, workers, and coordinators. Leaders manage UAVs, workers conduct ML model training, and coordinators facilitate data relaying between IoT devices and worker UAVs. They designed a novel methodology called Hierarchical Nested Personalized Federated Learning (HN-PFL), which exploits meta-gradient that captures data commonalities across the disconnected device clusters and yield personalized local models [38]. It introduces a two-layer hierarchical structure, involving swarm-level (local) aggregations within UAV swarms and global-level aggregations facilitated by access points (APs) or the core network. Zhong et al. also employed hierarchical over-the-air aggregation, utilizing a UAV as a Parameter Server (UAV-PS) (UAV leader) for collaborative ML training across widely distributed devices. The UAV-PS flies across its large service area to serve more devices. After the UAV-PS completes one round, it further aggregates the received partially aggregated local gradients to obtain a noisy version of the desired global gradient for global model updating [39]. Moreover, the researchers proposed a method to fine-tune the UAV's flight path and the way data is aggregated, aiming to achieve the best possible model performance as measured by the Mean Squared Error (MSE). They employed a specific algorithmic approach, combining Alternating Optimization (AO) and Successive Convex Approximation (SCA), to efficiently solve this complex optimization problem.

Ruby et al. introduced a two-tier FL network where IoT devices serve as core data holders, low altitude aerial platforms (UAVs) act as mid-tier model aggregators, and high-altitude UAVs function as the top-tier model aggregator [32]. The study focuses on addressing the energy-efficient computation and communication resource allocation challenge. Key contributions include adopting dual decomposition techniques to solve computation and communication resource allocation problems, proposing client-edge assignment schemes considering both energy consumption and client importance, and conducting extensive simulations to validate the effectiveness of the proposed scheme. The results emphasize the importance of considering not only energy consumption but also data distribution in client-edge assignments in two-tier hierarchical FL networks.

5.1.4 Aggregation algorithm

Aggregation in non-IID data involves combining locally trained models from different devices to create a global model. This process must be carefully designed to address the disparities in data characteristics among devices. A common aggregation technique in FL is where models from all devices contribute equally to the global model. While simple, FedAVG may not be optimal for non-IID scenarios. Wang et al. proposed a FedAVG algorithm for non-IID, named AGI-fedavg [44]. AGI-fedavg enhances the FedAvg algorithm by incorporating the data characteristics of each device. In each iteration of the FL process after local training on devices, each device returns the locally updated model weights and the data characteristics owned by each device to the server (i.e., data length, label). As each device returns its model parameters and labels, node eigenvalues are calculated based on the FedAvg algorithm during the weight computation. Finally, the calculated node eigenvalue and global weight are calculated, and the weight value is sent back to the device according to the corresponding label of each device for the next round of iteration. This algorithm allows the server to consider not only the model parameters but also contextual information about the data, enhancing the FL process in the presence of non-IID data [44].

A new aggregation algorithm named FedBA, introduced by Li et al., addresses the non-IID issue in a UAV-assisted FL framework. The FedBA algorithm incorporates a Euclidean distance function to evaluate the difference between two models: the global model from the previous communication round and each local device's model. The resulting values, after normalization, are then used to calculate aggregation weights for obtaining the global model for the new round. This algorithm ensures that each device's model contributes to the global model based on its dissimilarity [35].

Yao and Cao used the FedAvg-Adam algorithm for aggregation and proposed an enhanced Convolutional Neural Network (CNN) by incorporating an additional batch normalization (BN) layer resulting a faster convergence speed in the presence of non-IID data [30].

Puppala et al. proposed an FL-based architecture to address the challenge of detecting contraband in X-ray images used for airport luggage security while prioritizing user privacy [45]. They focus on the airport IoT environment and employ the FedAvg algorithm adapting it to suit the unique characteristics of the airport setting. The FedAvg algorithm is found to be robust against non-IID data, reducing the number of communication links needed for training deep networks on decentralized data.

Amit and Mohan presented an updated version of FedAvg, addressing dataset diversity using transfer learning and dynamic weight allocation [41]. Transfer learning involves using a pre-trained model and fine-tuning its weights with incremental adjustments to the learning rate in the new model. This method helps to extract the features through domain adaptation at higher layers. Khullar and Singh used incremental learning to the continuous improvement of the model's performance over time as it receives new data [42].

Two approaches, FedEx-Sync and FedEx-Async, are proposed and evaluated by Bian et al. [47]. FedEx-Sync adopts a synchronous learning approach, where all devices synchronize their learning rounds, the waiting time allows all devices to align their learning rounds, contributing to a synchronous update of the global model. On the other hand, FedEx-Async follows an asynchronous approach, allowing devices to initiate new rounds immediately and update the global model without predefined waiting periods. The performance comparison indicates that FedEx-Sync's performance is comparable to FedEx-Async when employing the Dirichlet distribution method, while FedEx-Sync outperform FedEx-Async under the simulation of a geographically-dependent distribution. Li et al. proposed an opportunistic and proactive transmission scheme to address the challenge of asynchronous model updates in FL [19]. In this scheme, each device is allowed to upload intermediate model updates to the server opportunistically, taking into account the wireless channel conditions. This scheme is advantageous in the non-IID context since it penalizes the local model from overfitting the biased local dataset.

Other authors look at how to best select devices for aggregation. They use algorithms to pick devices that improve the global model performance when there is diversity in the data. Zhang et al. proposed the FedFreq algorithm. They observed that the parameter distribution of the global model will be biased towards devices that often participate in FL, which is not friendly to the robustness of the global model. Therefore, their insight is to reduce the impact of devices used with high training frequency on the global model to improve the robustness of the model [24]. Cheriguene et al. used the Structural Similarity Index Measure (SSIM) to compute the dissimilarity of data between clients. SSIM is made up of three components, namely the visual impact of changes in image brightness, contrast, and any remaining defects, together known as structural alterations. The selection process in the study [36] prioritizes devices with the highest SSIM scores for participation in the FL round due to their perceived dissimilarity in data. Deng et al. proposed an enterprise-oriented framework to find FL devices with similar data resources. The authors in this study [31] employed a domain adaptation method within the context of transfer learning. This method was utilized to extract domain-invariant features, mitigating the impact of data differences among FL devices.

5.1.5 Optimization algorithms

Mashhadi et al. proposed a trajectory optimization algorithm for drones based on collaborative training between intelligent wireless devices and drones. The drone is interconnected with ground-distributed devices and shares neural network parameters. Ground-intelligent devices collect data locally and train the network, then transmit and aggregate parameters with drones, ultimately achieving convergence. The optimization considers various factors, such as convergence rate, communication errors, sensor noise levels, and the characteristics of heterogeneous local datasets. Experimental results across different datasets scenarios (IID and Non-IID) demonstrate the algorithm's higher accuracy [33].

The Yao and Sun investigated the CPU frequency optimization problem in an Internet of Drones (IoD) network. The goal was to minimize the energy consumption of all the drones during the FL training process while satisfying the latency requirement of FL training time. The study [37] introduces an algorithm with polynomial time complexity to determine the optimal solution. The evaluation indicates that the proposed algorithm consumes less energy when dealing with non-IID data compared to a baseline algorithm. This is beneficial because, in scenarios with non-IID data, the local training times of different drones vary significantly. Consequently, the FL process is more likely to be affected by the slowest drone, which becomes the bottleneck in the non-IID case.

Donevski et al. proposed an FL approach to enhance the performance of autonomous road vehicles using a Drone Traffic Monitor (DTM). The primary focus in this study [27] is on quickly learning a specific critical object (CO) class, considering non-IID data across devices with varying computational capabilities. The proposed solution involves dynamic resource allocation based on each device's contribution and incorporates heuristic measures such as maximizing or equalizing epochs computed across learners. The experiments utilize the FedProx FL algorithm in computer-vision tasks, demonstrating the effectiveness of the solution in improving system accuracy and rapidly learning the underrepresented CO class.

5.1.6 Federated Learning Architecture

A novel joint FL framework is introduced by Yu et al., addressing the challenges of handling hybrid vertically and horizontally partitioned data. Vertical FL involves different features for the same samples distributed across devices, whereas horizontal FL involves the same features but with different samples of data. The proposed framework allows cooperative training between the server and devices, where local models are trained independently and then aggregated to form a global model. The experimental results demonstrate that this joint FL framework achieves rapid convergence [43].

Qu et al. proposed a decentralized FL architecture for UAV networks (DFL-UN) in which each UAV client engages in both, local model training and aggregation of models from neighboring UAV clients. This eliminates the need for a server for global model aggregation. The DFL-UN demonstrates effectiveness in achieving comparable learning performance with reduced training latency in the presence of non-IID data [2].

The Liu et al. presented a framework called Intermittent FL designed to capture the realistic challenges posed by intermittent communication outages in cellular-connected UAV networks. The authors [28] evaluate the impact of communication outages on the learning accuracy, considering scenarios with both IID and non-IID datasets.

He et al. introduced a collaborative intrusion detection algorithm, utilizing a conditional generative adversarial net (CGAN) to tackle the issue of small samples and data imbalance. They apply LSTM networks in the generator and discriminator of CGAN to retain the contextual information for a long time and to portray small variations between normal and attack data. The evaluation demonstrates the efficiency of the algorithm in addressing the dataset imbalance [34].

5.2 Client–server communication

This section categorizes how UAV clients exchange updated parameters during the aggregation step, distinguishing between synchronous and asynchronous communication [64]. In synchronous aggregation, model aggregation occurs only after all UAV client updates have reached the UAV leader. This ensures that all UAV clients synchronize their progress before the aggregation step. On the other hand, the primary goal of asynchronous aggregation is to accelerate the training process. In asynchronous aggregation, the UAV leader aggregates the updated parameters as soon as it receives local updates from each UAV client. This allows each UAV client to train independently without waiting for others to complete their updates. In practical applications, selecting the appropriate type of communication UAV client is crucial since it has a significant impact on the efficiency of the FL architecture.

A review of studies on FL reveals that a substantial portion of the study has focused on using the synchronized type of exchange model parameters among UAV clients. However, several studies have chosen asynchronous communication, which offers distinct advantages in certain scenarios. The study by the authors of [19] focuses on mitigating staleness in the global model resulting from asynchronous aggregation. This issue becomes particularly pronounced in scenarios where a UAV faces dynamic wireless transmission challenges. The study of Bian et al. employs both synchronous and asynchronous methods, depending on the situation; if a UAV client with a shorter processing time needs to wait for one with a longer processing time, asynchronous communication is preferred, and vice versa. Similarly, Donevski et al. supports the use of asynchronous communication to enhance learning efficiency in heterogeneous networks with diverse computational capacities and limited resources. Additionally, the study of Qu et al. designs the training process to be asynchronous and fully distributed across multiple UAVs, highlighting the highly dynamic and potentially unstable nature of UAV networks.

5.3 Statistical heterogeneity in federated learning

This analysis delves into the various methodologies that confront the challenges posed by statistical heterogeneity within federated learning frameworks. Statistical heterogeneity, characterized by the diverse distributions of data across client devices, can profoundly affect model training and overall predictive performance.

5.3.1 Client clustering

By grouping clients that exhibit similar data distributions, client clustering methods are designed to tackle the challenges posed by statistical heterogeneity in federated learning environments. These methods ensure that the training process becomes more streamlined and effective. To determine the similarity between clients, various approaches utilize specific metrics, such as cosine similarity or Euclidean distance, which are calculated based on model updates or gradient vectors. This methodology is particularly advantageous as it directly mitigates the issue of statistical heterogeneity, preventing clients with fundamentally different data distributions from undermining each other's model updates and overall performance.

A prime example of this approach is FedBA [35], which specifically addresses non-IID data in UAV networks. FedBA leverages Bayesian aggregation techniques to identify and group UAVs that share similar data characteristics. By doing so, it forms more homogeneous sub-federations, allowing these clusters to collaborate more effectively in training, ultimately leading to enhanced model performance across the network.

5.3.2 Hierarchical federated learning

These methodologies leverage multi-tiered architectures specifically designed to manage heterogeneity in data sources and processing. By implementing intermediary aggregation layers, the approach effectively mitigates the challenges posed by diverse local data characteristics. This design allows for the development of specialized sub-models tailored to different hierarchical levels, thereby facilitating partial personalization while ensuring that global knowledge is still shared across the system.

In terms of architectural design, the system typically comprises three distinct tiers. Lower-tier aggregations focus on local aggregations among similar devices, optimizing data processing by grouping together devices that exhibit comparable characteristics or data patterns. In middle-tier aggregations, data is aggregated at edge servers or UAVs. This tier acts as a critical bridge, enabling efficient data processing closer to the source and reducing latency, all while managing the complexity of integrating various data streams. Upper-tier Aggregation involves gathering data at a central server, which synthesizes information from the lower and middle tiers. This tier consolidates comprehensive insights, enabling large-scale data analysis and decision-making that supports a broader application of the shared knowledge.

An illustrative example of this approach can be seen in a hierarchical nested method, which assigns UAVs the responsibility of coordinating training among ground devices that exhibit comparable characteristics [38]. This structure fosters a multi-level personalization strategy that effectively addresses the issue of heterogeneity at each tier of the network.

Another pertinent example is an edge-aided framework where edge nodes function as intermediary facilitators between clients and the central server [5]. Clients that possess similar data distributions are connected to the same edge node, which performs localized aggregations to mitigate the effects of data heterogeneity. The central server subsequently aggregates these pre-processed and smoothed models derived from the edge nodes, thus enhancing the overall efficiency and effectiveness of the data processing pipeline.

5.3.3 Hybrid and split learning

The hybrid or split learning approach optimizes data privacy and computational efficiency by partitioning the neural network into distinct segments. Initially, early layers are processed on client devices, ensuring that raw data remains on the device. Only the intermediate feature representations-outputs from these layers-are sent to a central server. At the server, the remaining layers of the neural network are computed using these representations, which allows for model training without accessing sensitive data. Federated aggregation techniques are then applied to merge model updates from various clients, effectively integrating knowledge while maintaining privacy and security in the learning process.

Split learning methods decrease communication demands, which is crucial when statistical variation would otherwise necessitate additional communication rounds for convergence. By reducing the dimensionality of transmitted data-specifically, sending relevant features instead of complete gradients, statistical heterogeneity can be effectively tackled across datasets. This approach streamlines communication and enhances the clarity of shared information. Standardizing intermediate representations also normalizes differences in data distributions, reducing heterogeneity and improving model consistency. Furthermore, enabling the server to perform complex computations on these standardized features optimizes resource use, leading to more efficient and accurate model training.

The split learning-assisted multi-UAV system for image classification utilizes feature extraction on UAVs, followed by processing at the server [29]. This approach helps standardize heterogeneous image data collected from various environments.

5.3.4 Specialized aggregation techniques

These methodologies enhance how client updates are evaluated and integrated. Instead of simple averaging of model updates, they utilize different functionalities. Adaptive Weighting Schemes adjust the influence of client updates based on the characteristics of their data, considering factors like volume and quality. Tailored loss functions address variations in data distributions, improving the model's ability to learn from diverse datasets and reducing the impact of unbalanced data. Regularization Techniques introduce constraints that prevent overfitting to specific data distributions, promoting generalization and robustness across varied datasets. In combination, these approaches lead to a more effective and resilient learning framework in distributed environments.

For instance, AGI-Fedavg addresses data heterogeneity in federated learning through a targeted approach [44]. It starts by evaluating client contributions based on quality rather than merely data volume, ensuring that the most relevant data influences the model effectively. Additionally, AGI-Fedavg adjusts aggregation weights according to specific performance metrics, prioritizing contributions from clients that enhance overall model performance.

Further, a semi-supervised approach by Zhang et al. uses unlabeled data to improve robustness across heterogeneous drone-captured images. The approach utilized advanced self-supervised learning techniques to extract generalizable features that perform well across diverse data distributions. It also employed consistency regularization to ensure stable model predictions by comparing labeled and unlabeled samples, enhancing the robustness and adaptability of the learning framework.

Table 3 Overview of UAV use cases utilizing FL to simulate non-IID data

Use cases	References
Surveillance	Military surveillance [19]
	Distributed surveillance in smart cities [38]
	Airport surveillance activities [41]
	Traffic monitoring [34]
Collaborative learning	UAVs collaboratively train ML model over a specific area. [20]
	Various firms' UAVs capture images at low altitudes from diverse locations. [35]
	Serve devices distributed in a relatively large area [39]
	Smart sensing in remote areas with no communication infrastructure [47, 65]
Environmental monitoring	Fire tracking and flood monitoring [21]
	Disaster image classification [42]
	UAVs provide timely warnings for critical objects affecting autonomous vehicles. [27]
Recognition and classification tasks	Image recognition [22, 24, 33]
	Identify specific objects such as airports, factors and parking lots [29]
	Exploration tasks, terrain discrimination and classification [30]
Inspection tasks	Using UAVs to realize line patrol in multiple areas [44]
	Using UAV to detect power grid abrasion [46]
Security	Automated detection methods to check luggage for dangerous items in airport [45]
	Detect the type/model of the UAV using the transmitted RF signals [26]
Manufacturing industry	Planing machining parameters for aircraft structural parts [31]
Production	Agricultural production and farming [43]

5.3.5 Energy and resource optimization

These methodologies address both statistical and system heterogeneity while optimizing resource allocation. This involves selecting clients that enhance model diversity and maintain resource efficiency by evaluating their unique data contributions. Communication is adjusted based on the importance of client data and their device battery levels, prioritizing critical updates to conserve energy and bandwidth. By analyzing data characteristics, adaptive compression is applied, balancing bandwidth usage with the preservation of essential information for effective model training.

Liang et al. introduce an energy-aware scheduling framework for IoT applications that formalizes client selection as a multi-objective optimization problem. This approach effectively accounts for data heterogeneity by recognizing the varying quality and types of data from different IoT devices. Additionally, it incorporates energy constraints to ensure the operational longevity of devices. Advanced reinforcement learning techniques enable the system to adaptively balance model quality and energy consumption, dynamically refining its strategies based on real-time feedback to optimize scheduling decisions.

6 Applications

This section delves into the diverse use cases where UAVs, coupled with FL, play a pivotal role in enhancing efficiency, accuracy, and privacy, in the presence of non-IID data.

6.1 Use cases

This section delves into the diverse use cases where UAVs, coupled with FL, play a pivotal role in enhancing efficiency, accuracy, and privacy, particularly in scenarios involving non-IID data. As shown in Table 3, we have categorized the use cases derived from the included studies into general categories. These categories are not mutually exclusive but have been outlined based on the primary objectives or targets of each use case. This classification aids in gaining a clearer understanding of the varied roles UAVs play in different applications, especially when integrated with FL. They are as follows:

Surveillance

Focuses on continuous observation and monitoring, encompassing a variety of applications. This includes employing UAVs as aerial users in military contexts, where they fly around target areas to collect data and support military applications through wireless networks [19]. In the context of distributed surveillance in smart cities, multiple UAVs are deployed throughout both urban and rural areas gathering data from an array of sensors and cameras. Another significant use case is airport surveillance [41] which utilizes a range of remote sensing technologies including security cameras, UAVs with onboard cameras, and specialized remote sensing cameras. Additionally, this surveillance category extends to traffic monitoring [34], showcasing the broad utility of UAVs in various surveillance scenarios.

Collaborative learning

Involves UAVs working together to train ML models over specific locations. Examples include UAVs from various firms capturing images at low altitudes in diverse locations, where each UAV employs its computational capabilities to perform local training on the data it has gathered [20, 35]. This approach also extends to serving devices distributed across relatively large areas [39] and enabling smart sensing in remote areas that lack communication infrastructure [47, 65].

Environmental monitoring

UAVs are utilized to monitor natural phenomena and environmental conditions. This includes fire tracking and flood monitoring [21] where a group of UAVs fly over a target area under server control to collect image data with equipped cameras. Each UAV, carrying a powerful processing unit (e.g., NVIDIA Jetson [10]), observing partial information of the target area. Another use case is in disaster image classification [42] which involves proposing and analyzing UAV-based disaster area image classification. Additionally, UAVs provide timely warnings for critical objects that could impact autonomous vehicles [27].

Recognition and classification tasks

Involve UAVs in image recognition and the identification of specific objects [22, 24, 33]. UAVs are particularly useful in detecting and classifying distinct objects or activities within complex environments such as airports and factories [29]. They also play a crucial role in exploration tasks, including terrain discrimination and classification [30].

Inspection tasks

Involve using UAVs for inspection tasks, such as patrolling transmission lines [44]. Since these lines are often in remote or harsh environments, manual inspection poses safety risks and inefficiencies. UAVs offer a safer, more efficient alternative for these inspections. Another use case is using UAVs for the detection of power grid abrasions [46] where UAVs primarily focus on identifying faults in the power grid, facilitating timely repair and maintenance.

Security

Focuses on enhancing safety and security. One use case is the use of automated detection methods in airports for checking luggage for dangerous items [45]. This involves collaboration among three airports in different countries to develop a global model for identifying contraband, despite challenges in sharing X-ray data due to international sensitivities. Another use case is using UAVs to detect the type/model of other UAVs by analyzing their radio frequency (RF) signals [26]. This method combines passive and active techniques and aims to identify the manufacturer of the UAV and model based on unique, custom-designed RF waveforms.

Manufacturing industry

UAVs are applied for various tasks such as planning and machining. A specific example is the planning machining parameters for aircraft structural parts [31].

Agricultural production

UAVs play a role in agricultural activities and farming [43]. They are used to collect farm data including crop growth prediction and pest diagnosis. UAVs gather varied samples from different farm areas to provide comprehensive agricultural insights.

7 Evaluation metrics

This examination focuses on the evaluation metrics that assess the performance and efficiency of FL within UAV applications, particularly in scenarios characterized by non-IID data distributions. These metrics are instrumental in measuring not only the accuracy and reliability of FL algorithms but also their scalability and resilience to heterogeneous data. By employing these metrics, the effectiveness of FL implementations can be evaluated in the face of the complexities presented by non-IID data environments.

7.1 Accuracy-based metrics

Accuracy is a critical evaluation metric across various domains, including FL applications for UAVs. It measures the proximity of predicted outcomes to their actual values, playing an essential role in assessing the effectiveness of ML models. A higher level of accuracy not only indicates more reliable and precise predictions but also emphasizes the overall dependability of a model in practical applications.

This metric has been extensively referenced in numerous studies, demonstrating its importance in evaluating the performance of federated learning models. In particular, when applied to image classification tasks involving UAVs, accuracy serves as a clear benchmark for how effectively these models can interpret and analyze visual data captured from aerial viewpoints. The focus on accuracy in this setting is vital, as it directly influences the operational success and efficiency of UAV applications across various industries, from agriculture to surveillance [23, 24, 26, 30, 35, 44].

[10] https://www.nvidia.com/en-us/autonomous-machines/embedded-systems/

Furthermore, the **Intersection over Union (IoU)**, as utilized in [33], measures the accuracy of the overlap between predicted and actual image regions. **Precision** and **Recall**, as detailed in [34], measure the accuracy of positive predictions and the completeness of true positive detection, respectively. The **F1 Score**, which represents the harmonic mean of precision and recall, is employed in [31, 34]. Additionally, model performance is evaluated with a particular emphasis on loss functions during both training and testing phases, as noted in [38]. The **Standard Mean Average Precision (mAP)**, used to assess the area under the precision-recall curve, is applied in [41].

7.2 Communication & resource efficiency metrics

Communication overhead refers to the measurable effects of data transfers on both **bandwidth consumption** and **transmission frequency** within a system. In the context of FL applications designed for UAVs, various evaluation metrics are employed to thoroughly assess technical model performance. A primary focus is on enhancing communication efficiency, which seeks to minimize the amount of data exchanged between edge devices. Implementing effective strategies in this area can significantly reduce latency and improve the overall responsiveness of the system [19, 21].

In addition to communication efficiency, **resource utilization** is another crucial metric that evaluates the effectiveness with which edge devices deploy their resources during operation. This encompasses factors such as CPU usage, memory consumption, and energy efficiency, all of which are vital for the sustainable operation of UAVs. A comprehensive examination of this topic discusses various methodologies aimed at optimizing resource allocation in edge devices [42]. The goal is to strike a balance between performance and resource constraints while ensuring privacy and security are maintained.

Finally, **energy efficiency** is essential in optimizing UAV-enabled FL systems. This involves implementing energy-aware participant selection methods to identify devices based on their energy resources and computational abilities, thereby enhancing system sustainability. Additionally, adopting energy-efficient strategies is crucial for minimizing energy consumption during model training and inference [37, 38].

7.3 Convergence analysis

Studies commonly report accuracy alongside other metrics to provide a comprehensive understanding of the overall performance of the model. Convergence analysis is a critical aspect in evaluating the performance of FL models for UAV applications. It focuses on understanding the behavior of the ML model and performance as it iteratively refines its parameters through collaboration with distributed UAV clients. **Convergence** refers to the point at which the ML model reaches stability or achieves a desired level of performance [35, 39]. A higher convergence rate indicates that the FL model approaches the optimal solution more rapidly during the training process.

Furthermore, **training latency** refers to the time it takes from starting the training of a machine learning model to receiving the results. It involves data preprocessing, model setup, and hardware performance. Factors such as model complexity and dataset size affect this delay, making it important to minimize for efficient real-time applications [2, 25].

8 Open challenges and future directions

In this section, we discuss potential research directions and highlight unresolved issues associated with FL for UAV applications in the presence of statistical heterogeneity.

Use of realistic datasets: FL relies on training ML models across decentralized devices. In the context of UAV applications, the datasets used to train these models play a pivotal role in determining the models' performance and applicability to real-world scenarios. However, many studies use non-realistic datasets for convenience and computational efficiency, which might not accurately represent the complexities and variations encountered by UAVs in actual operational environments. An example of this is using the MNIST [62] dataset to simulate FL in UAV applications [5]. Future efforts should focus on integrating realism data into the training process to ensure that the trained models are more robust, adaptable, and capable of addressing the challenges encountered in practical, real-world scenarios. This is crucial, as pointed in [24, 44], since the authors need to evaluate their algorithm in a real environment to demonstrate its practical efficacy and adaptability.

Training data distributions: This aspect emphasizes the significance of simulating diverse data distributions during FL experiments for UAV applications. Each UAV client collects data with varying characteristics, features, labels, sizes, or a mix of these aspects. Notably, the discussed studies have mostly simulated label partitioning, often with distributions closely resembling IID data distributions (cf Table 2). However, there is a gap in exploring non-IID distributions that better reflect the real-world applications. Future studies should focus on simulating various types of distributions to ensure a more accurate representation of the complexities encountered in UAV practical scenarios. It is critical to add more variability in the dataset, as noted in [41]. Using mixed-type distributions, which includes variations in features, labels, and sizes, is vital to enhance the realism of FL experiments in UAV applications. Moreover, as pointed out in [44], it is suggested that incorporating additional UAV clients for training in future research would be more effective in simulating non-IID data in real-world environments. This suggest enables the global model to learn from a broader spectrum of real-world scenarios, which is crucial for applications requiring high accuracy and adaptability in varying conditions.

Catastrophic forgetting: FL methods included in this review are designed within the framework of static UAV application scenarios, where the training data are predetermined and assumed to remain fixed. This approach contrasts with the dynamic nature of real-world applications, where new data are generated regularly and its distribution can change significantly. A major challenge arises when an ML model is confronted with new data that significantly differs from the data previously used for training, exposing UAV clients to the problem of catastrophic forgetting [66]. Catastrophic forgetting is a phenomenon in ML where a model forgets previously learned information upon learning new data. This issue is significant in scenarios where models are continuously trained on new data streams. The model, while adapting to the new information, tends to completely overwrite or lose the knowledge it had acquired from the older data. Continual learning [67] is employed in this context to help the ML model adapt to new data. This technique should be explored further in this domain, particularly in non-IID data scenarios with new training data distributions. Therefore, it is crucial for researchers to develop solutions that can better adapt to dynamic real-world scenarios.

Limitation and future perspective: A key limitation identified in the systematic review is the need for FL frameworks to adapt to non-IID distributions of training data and the dynamic nature of UAV application environments [27]. This adaptation is essential for ensuring the effectiveness of FL in real-world UAV scenarios. The authors' future direction [24] involves enhancing their proposed Semi-Supervised Federated Learning (SSFL) algorithm to maximize the use of unlabeled data and further develop the underlying theory, enabling more effective applications in real-world scenarios. Additionally, other authors [41] suggest that the FL algorithm they have proposed could be further evaluated in various applications, such as image segmentation and image enhancement tasks.

The integration of multi-model [68] support within the FL training process, represents a significant advancement. This development would be enabled by separating the global model aggregation from local training processes. Such a separation would grant UAV clients the flexibility to implement various learning algorithms as per their specific needs.

9 Summary

Federated Learning (FL) presents itself as a promising solution for training Machine Learning (ML) models with large and diverse datasets, without compromising information confidentiality. This characteristic is significant for UAV applications, where UAV data is inherently sensitive to privacy and often cannot be easily shared. In this review, we emphasize a critical challenge in FL in the context of UAV applications, specifically focusing on non-IID (Non-Independently and Identically Distributed) distributions of data. We surveyed and classified 31 studies published between 2016 and October of 2023 to address five research questions. We introduced the challenges posed by non-IID data in UAV applications, emphasizing their impact on ML model performance, including learning accuracy, stability of the FL algorithm, convergence behavior, and communication efficiency. We delved into the training data distributions, focusing on quantity, label, feature, and mixed-type distributions. We noted that label skew was a common distribution type used. The most commonly unrealistic dataset employed is MNIST, which did not effectively simulate real-world UAV data, given the differences in the data domain, size, and complexity.

We systematically reviewed state-of-the-art FL frameworks designed to address non-IID data, including techniques like data sharing, split learning, clustering, FL architecture, aggregation algorithm and optimization algorithms.

Additionally, we discussed and highlighted use cases of FL on non-IID data in areas like surveillance, collaborative learning, environmental monitoring, recognition and classification tasks, inspection, security, manufacturing, and production. We shed light on the evaluation metrics employed, accuracy, and convergence rate, highlighting their significance in this field.

The comprehensive systematic literature review presented in this study is expected to guide researchers in understanding the state-of-the-art and inform future studies on FL with non-IID data.

Appendix: Summarized table of federated learning studies in UAV applications

See Tables 4 and 5.

Table 4 Federated learning studies for UAV applications

FL framework	Contribution
OPT-HSFL [19]	Aims to mitigate the impact of dynamic wireless conditions on model transmission
HN-PFL [38]	Exploits meta-gradient based learning across disconnected device clusters and yield personalized local models
clustered decentralized FL [20]	Divides the UAV network into clusters in iterative way for local model aggregation while ensuring connectivity among Cluster head UAVs
HSFL algorithm [21]	Encompasses the parallel model training mechanism of FL and the model splitting structure of Split Learning
[33]	Optimize the drone trajectory to achieve the fastest learning and the best final performance for the trained NN model
CGAN [34]	Collaborative intrusion detection algorithm based on CGAN-LSTM with blockchain empowered distributed FL
FedBA [35]	Alleviate the problem of data heterogeneity in UAV-assisted FL
C-FLA [23]	Two optimization strategies for handling multiple classifiers, which effectively address the challenges posed by client heterogeneity
OA-FL system [39]	A system which using UAVs as a server to aggregate local gradients hierarchically in large areas, addressing the challenges of communication and straggler issues
[42]	Disaster image classification in the context of the Internet of UAVs
Hierarchical FL algorithm [5]	A high-performing FL scheme for the edge-aided UAV network that works well in real-world scenarios with non-IID distributions (i.e., highly skewed feature and label distributions)
SSFL Framework [24]	FL framework for enhanced data privacy, developing a robust semi-supervised FL system, proposing a novel model aggregation rule to handle statistical heterogeneity
TT-HF [25]	Efficiency of FL in D2D-enabled wireless networks by augmenting global aggregations with cooperative consensus procedure among device clusters
DEEPS [36]	Participant selection scheme that prioritizes participants with high data diversity and sufficient battery capacity to handle local training
[29]	Split Learning assisted multi UAV system for image classification tasks in area exploration scenarios
AGI-Fedavg [44]	An FL algorithm for power grid data, addressing privacy and non-IID challenges
[45]	FL-based architecture to detect contraband in x-ray baggage security images while maintaining user privacy
[30]	Land classification method based on FL which uses Fedavg-Adam algorithm and introduce an improved CNN
[46]	Adaptive method according to the idea of dynamic adjustment of static parameters such as learning rate and gradient
[37]	CPU frequency optimization in an Internet of Drone network to reduce energy use during FL training while meeting latency requirements
An enterprise-oriented framework [31]	Framework to find FL participants with similar data resources while minimizing the disclosure of enterprise information
FedEx [47]	FL framework use in situations where direct communication between the server and clients is not possible
[32]	Framework and resource allocation strategies for energy-efficient FL in a two-tier network with IoT devices, UAV aggregators, and consideration of non-IID data
HSFL [22]	Improves communication efficiency and learning accuracy under non-IID data distributions along with a MAB-based user selection scheme
Joint federated learning [43]	FL framework for Edge-assisted Internet of Agriculture Things, coupled with a resource-constrained device scheduling algorithm, to enhance convergence, communication efficiency, and model accuracy
[41]	FL-based data management framework for airport object representations that enhances security and privacy by preserving data on the client side while achieving better detection accuracy and communication efficiency in object detection tasks
[26]	Deep learning method to detect the type/model of the UAV using the transmitted RF signals
[27]	FL in the context of autonomous traffic monitoring with a drone orchestrator and ground-based learners to enhance learning in a dynamic and non-IID data environment
DFL-UN [2]	FL architecture called Decentralized FL for UAV Networks, which enables FL within UAV networks without a central server
[28]	An intermittent FL model that accounts for uplink communication outages which develops a tractable approach to analyze and characterize the uplink outage probability

Table 4 (continued)

FL framework	Contribution
H-Home [65]	Framework that combines FL and RL to address offload management challenges in Flying Ad-Hoc Networks

Table 5 List of abbreviations and synonyms used in our systematic review

Abbreviation	Full label	Synonym(s)
ML	Machine learning	
CNN	Convolutional neural network	
DDRNet	Deep dual-resolution network	
FCNN	Fully connected neural network	
RNN	Recurrent neural network	
RF	Random forest	
DT	Decision tree	
SVM	Support vector machine	
GAN-LSTM	Generative adversarial network with long short-term memory	
CGAN-LSTM	Conditional generative adversarial network with long short-term memory	
FL-CGAN-LSTM	Federated learning conditional generative adversarial network with long short-term memory	
VGG16	Visual geometry group 16	
ResNet152	Residual network 152	
Inception ResNet	Inception residual network	
Xception	Extreme inception	
ResNet-18	Residual network 18	
R-CNN	Region-based convolutional neural network	
YOLOV3	You only look once version 3	
YOLOv7	You only look once version 7	
SSD	Single shot MultiBox detector	
CNN with BN	Convolutional neural network with batch normalization	
RL	Reinforcement learning	
ResNet-9	Residual network 9	
AlexNet	Alex neural network	
FL	Federated learning	
IID	Independently and identically distributed	
Non-IID	Non-independently and identically distributed	Statistical heterogeneity
IOT devices	Internet of thing devices	Ground-intelligent devices
UAV leader	Unmanned aerial vehicle leader	Base station (BS), server, unmanned aerial vehicle Parameter server (UAV_PS), Cluster-head (CH) UAV
UAV client	Unmanned aerial vehicle client	Device, drone

Acknowledgements The authors gratefully acknowledge the Database and Information Systems group at the Department of Computer Science and Automation, Technical University Ilmenau, for their valuable support throughout this research.

Funding Open Access funding enabled and organized by Projekt DEAL.

Data availability Not applicable.

Declarations

Conflict of interest No funds, grants, or other support were received, and the authors have no relevant financial or non-financial interests to disclose.

References

1. Brik, B., Ksentini, A., Bouaziz, M.: Federated Learning for UAVs-Enabled Wireless Networks: Use Cases, Challenges, and Open Problems. IEEE Access **8**, 53841–53849 (2020). https://doi.org/10.1109/ACCESS.2020.2981430
2. Qu, Y., Dai, H., Zhuang, Y., Chen, J., Dong, C., Wu, F., Guo, S.: Decentralized Federated Learning for UAV Networks: Architecture, Challenges, and Opportunities. IEEE Netw. **35**(6), 156–162 (2021). https://doi.org/10.1109/MNET.001.2100253
3. Goudarzi, S., Kama, N., Anisi, M.H., Zeadally, S., Mumtaz, S.: Data collection using unmanned aerial vehicles for Internet of Things platforms. Comput. Electr. Eng. **75**, 1–15 (2019). https://doi.org/10.1016/j.compeleceng.2019.01.028
4. Bithas, P.S., Michailidis, E.T., Nomikos, N., Vouyioukas, D., Kanatas, A.G.: A Survey on Machine-Learning Techniques for UAV-Based Communications. Sensors **19**(23), 5170 (2019). https://doi.org/10.3390/s19235170
5. Tursunboev, J., Kang, Y.-S., Huh, S.-B., Lim, D.-W., Kang, J.-M., Jung, H.: Hierarchical Federated Learning for Edge-Aided Unmanned Aerial Vehicle Networks. Appl. Sci. **12**(2), 670 (2022). https://doi.org/10.3390/app12020670
6. Zhang, Z., Zhu, L.: A Review on Unmanned Aerial Vehicle Remote Sensing: Platforms, Sensors, Data Processing Methods, and Applications. Drones **7**(6), 398 (2023). https://doi.org/10.3390/drones7060398
7. Mekdad, Y., Aris, A., Babun, L., Fergougui, A.E., Conti, M., Lazzeretti, R., Uluagac, A.S.: A survey on security and privacy issues of UAVs. Comput. Netw. **224**, 109626 (2023). https://doi.org/10.1016/j.comnet.2023.109626
8. Alsamhi, S.H., Shvetsov, A.V., Kumar, S., Hassan, J., Alhartomi, M.A., Shvetsova, S.V., Sahal, R., Hawbani, A.: Computing in the Sky: A Survey on Intelligent Ubiquitous Computing for UAV-Assisted 6G Networks and Industry 4.0/5.0. Drones **6**(7), 177 (2022). https://doi.org/10.3390/drones6070177
9. Gupta, O., Raskar, R.: Distributed learning of deep neural network over multiple agents. J. Netw. Comput. Appl. **116**, 1–8 (2018). https://doi.org/10.1016/j.jnca.2018.05.003
10. Wen, J., Zhang, Z., Lan, Y., Cui, Z., Cai, J., Zhang, W.: A survey on federated learning: challenges and applications. Int. J. Mach. Learn. Cyb. **14**(2), 513–535 (2023). https://doi.org/10.1007/s13042-022-01647-y
11. McMahan, H.B., Moore, E., Ramage, D., Hampson, S., Arcas, B.A.y.: Communication-Efficient Learning of Deep Networks from Decentralized Data (2016). https://arxiv.org/abs/1602.05629v4
12. Hard, A., Rao, K., Mathews, R., Ramaswamy, S., Beaufays, F., Augenstein, S., Eichner, H., Kiddon, C., Ramage, D.: Federated Learning for Mobile Keyboard Prediction (2018). https://arxiv.org/abs/1811.03604v2
13. Chen, Y., Wang, J., Yu, C., Gao, W., Qin, X.: FedHealth: A Federated Transfer Learning Framework for Wearable Healthcare (2019). https://arxiv.org/abs/1907.09173v2
14. Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., Shamseer, L., Tetzlaff, J.M., Akl, E.A., Brennan, S.E., Chou, R., Glanville, J., Grimshaw, J.M., Hróbjartsson, A., Lalu, M.M., Li, T., Loder, E.W., Mayo-Wilson, E., McDonald, S., McGuinness, L.A., Stewart, L.A., Thomas, J., Tricco, A.C., Welch, V.A., Whiting, P., Moher, D.: The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ, 71 (2021) https://doi.org/10.1136/bmj.n71
15. Li, T., Sahu, A.K., Talwalkar, A., Smith, V.: Federated Learning: Challenges, Methods, and Future Directions. IEEE Signal Process. Magaz. **37**(3), 50–60 (2020). https://doi.org/10.1109/MSP.2020.2975749
16. Yang, Z., Liu, Y., Zhang, S., Zhou, K.: Personalized federated learning with model interpolation among client clusters and its application in smart home. World Wide Web **26**(4), 2175–2200 (2023). https://doi.org/10.1007/s11280-022-01132-0
17. Zhu, H., Xu, J., Liu, S., Jin, Y.: Federated learning on non-iid data: A survey. Neurocomputing **465**, 371–390 (2021). https://doi.org/10.1016/j.neucom.2021.07.098
18. Haddaway, N.R., Page, M.J., Pritchard, C.C., McGuinness, L.A.: PRISMA2020: An R package and Shiny app for producing PRISMA 2020-compliant flow diagrams, with interactivity for optimised digital transparency and Open Synthesis. Campbell System. Rev. **18**(2), 1230 (2022). https://doi.org/10.1002/cl2.1230
19. Li, J., Liu, X., Mahmoodi, T.: Opportunistic Transmission of Distributed Learning Models in Mobile UAVs. arXiv. arXiv:2306.09484 [cs] (2023). doi:10.48550/arXiv.2306.09484
20. Hoang, D.N.M., Truong, V.T., Le, H.D., Le, L.B.: Clustered and Scalable Federated Learning Framework for UAV Swarms. TechRxiv (2023). https://doi.org/10.36227/techrxiv.22730012.v1
21. Liu, X., Deng, Y., Mahmoodi, T.: Wireless Distributed Learning: A New Hybrid Split and Federated Learning Approach. IEEE Trans. Wireless Commun. **22**(4), 2650–2665 (2023). https://doi.org/10.1109/TWC.2022.3213411
22. Liu, X., Deng, Y., Mahmoodi, T.: A Novel Hybrid Split and Federated Learning Architecture in Wireless UAV Networks. In: ICC 2022 - IEEE International Conference on Communications, pp. 1–6 (2022). https://doi.org/10.1109/ICC45855.2022.9838867. ISSN: 1938-1883
23. Pei, J., Yu, Z., Liu, W., Kong, W., Zhang, J., Wang, L.: Clustered federated learning multi-classifier for non-IID scenario in drone devices. In: Proceedings of the 5th International ACM Mobicom Workshop on Drone Assisted Wireless Communications For 5G and Beyond. DroneCom '22, pp. 79–84. Association for Computing Machinery, New York, NY, USA (2022). https://doi.org/10.1145/3555661.3560868
24. Zhang, Z., Ma, S., Yang, Z., Xiong, Z., Kang, J., Wu, Y., Zhang, K., Niyato, D.: Robust Semi-supervised Federated Learning for Images Automatic Recognition in Internet of Drones. arXiv. arXiv:2201.01230 [cs] (2022). https://doi.org/10.48550/arXiv.2201.01230
25. Lin, F.P.-C., Hosseinalipour, S., Azam, S.S., Brinton, C.G., Michelusi, N.: Semi-Decentralized Federated Learning With Cooperative D2D Local Model Aggregations. IEEE J. Sel. Areas Commun. **39**(12), 3851–3869 (2021). https://doi.org/10.1109/JSAC.2021.3118344
26. Reus-Muns, G., Chowdhury, K.R.: Classifying UAVs With Proprietary Waveforms via Preamble Feature Extraction and Federated Learning. IEEE Trans. Veh. Technol. **70**(7), 6279–6290 (2021). https://doi.org/10.1109/TVT.2021.3081049
27. Donevski, I., Nielsen, J.J., Popovski, P.: On Addressing Heterogeneity in Federated Learning for Autonomous Vehicles Connected to a Drone Orchestrator. arXiv. arXiv:2108.02712 [cs] (2021). https://doi.org/10.48550/arXiv.2108.02712
28. Liu, C.-H., Liang, D.-C., Gau, R.-H., Wei, L.: Modeling and Analysis of Intermittent Federated Learning Over Cellular-Connected UAV Networks. In: 2022 IEEE 95th Vehicular Technology

Conference: (VTC2022-Spring), pp. 1–6 (2022). https://doi.org/10.1109/VTC2022-Spring54318.2022.9860913. ISSN: 2577-2465
29. Sun, T., Wang, X., Umehira, M., Ji, Y.: Split Learning Assisted Multi-UAV System for Image Classification Task. In: 2023 IEEE 97th Vehicular Technology Conference (VTC2023-Spring), pp. 1–6 (2023). https://doi.org/10.1109/VTC2023-Spring57618.2023.10200607. ISSN: 2577-2465
30. Yao, K., Cao, X.: UAV Land Classification Method Based on Federated Learning. Academic Journal of Computing & Information Science **5**(7) (2022) https://doi.org/10.25236/AJCIS.2022.050712
31. Deng, T., Li, Y., Liu, X., Wang, L.: Federated learning-based collaborative manufacturing for complex parts. J. Intell. Manuf. **34**(7), 3025–3038 (2023). https://doi.org/10.1007/s10845-022-01968-3
32. Ruby, R., Yang, H., Figueiredo, F.A.P., Huynh-The, T., Wu, K.: Energy-Efficient Multiprocessor-Based Computation and Communication Resource Allocation in Two-Tier Federated Learning Networks. IEEE Internet Things J. **10**(7), 5689–5703 (2023). https://doi.org/10.1109/JIOT.2022.3153996
33. Mashhadi, M.B., Mahdavimoghadam, M., Tafazolli, R., Saad, W.: Collaborative Learning with a Drone Orchestrator. arXiv. arXiv:2303.02266 [cs, eess, math, stat] (2023). https://doi.org/10.48550/arXiv.2303.02266
34. He, X., Chen, Q., Tang, L., Wang, W., Liu, T.: CGAN-Based Collaborative Intrusion Detection for UAV Networks: A Blockchain-Empowered Distributed Federated Learning Approach. IEEE Internet Things J. **10**(1), 120–132 (2023). https://doi.org/10.1109/JIOT.2022.3200121
35. Li, P., Liu, Z., Chang, L., Peng, J., Wu, Y.: FedBA: Non-IID Federated Learning Framework in UAV Networks. arXiv. arXiv:2210.04699 [cs] (2022). https://doi.org/10.48550/arXiv.2210.04699
36. Cheriguene, Y., Jaafar, W., Kerrache, C.A., Yanikomeroglu, H., Bousbaa, F.Z., Lagraa, N.: Data-Efficient Energy-Aware Participant Selection for UAV-Enabled Federated Learning. arXiv. arXiv:2308.07273 [cs] (2023). doi:10.48550/arXiv.2308.07273
37. Yao, J., Sun, X.: Energy-Efficient Federated Learning in Internet of Drones Networks. In: 2023 IEEE 24th International Conference on High Performance Switching and Routing (HPSR), pp. 185–190 (2023). https://doi.org/10.1109/HPSR57248.2023.10147956. ISSN: 2325-5609
38. Wang, S., Hosseinalipour, S., Gorlatova, M., Brinton, C.G., Chiang, M.: UAV-Assisted Online Machine Learning Over Multi-Tiered Networks: A Hierarchical Nested Personalized Federated Learning Approach. IEEE Trans. Netw. Service Manage. **20**(2), 1847–1865 (2023). https://doi.org/10.1109/TNSM.2022.3216326
39. Zhong, X., Yuan, X., Yang, H., Zhong, C.: UAV-Assisted Hierarchical Aggregation for Over-the-Air Federated Learning. arXiv. arXiv:2209.11624 [eess] (2022). https://doi.org/10.48550/arXiv.2209.11624
40. Safa, A., Verbelen, T., Ocket, I., Bourdoux, A., Sahli, H., Catthoor, F., Gielen, G.G.E.: Learning to Encode Vision on the Fly in Unknown Environments: A Continual Learning SLAM Approach for Drones. In: 2022 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR), pp. 373–378 (2022). https://doi.org/10.1109/SSRR56537.2022.10018713. ISSN: 2475-8426
41. Amit, R.A., Mohan, C.K.: Federated Learning: Dataset Management for Airport Object Representations Using Remote Sensing Images. In: 2022 IEEE Aerospace Conference (AERO), pp. 1–14 (2022). https://doi.org/10.1109/AERO53065.2022.9843800. ISSN: 1095-323X
42. Khullar, V., Singh, H.P.: Privacy protected internet of unmanned aerial vehicles for disastrous site identification. Concurrency and Computation: Practice and Experience **34**(19) (2022) https://doi.org/10.1002/cpe.7040
43. Yu, C., Shen, S., Zhang, K., Zhao, H., Shi, Y.: Energy-Aware Device Scheduling for Joint Federated Learning in Edge-assisted Internet of Agriculture Things. In: 2022 IEEE Wireless Communications and Networking Conference (WCNC), pp. 1140–1145 (2022). https://doi.org/10.1109/WCNC51071.2022.9771547. ISSN: 1558-2612
44. Wang, M., Lei, Y., Liang, Y., Lv, X., Mo, W.: AGI-Fedavg: A UAV Power Line Inspection Algorithm Based on Federated Learning. In: 2021 IEEE 3rd International Conference on Civil Aviation Safety and Information Technology (ICCASIT), pp. 1030–1034 (2021). https://doi.org/10.1109/ICCASIT53235.2021.9633432
45. Puppala, S., Hossain, I., Talukder, S.: Towards Federated Learning Based Contraband Detection Within Airport Baggage X-Rays. In: 2022 IEEE International Conference on Machine Learning and Applied Network Technologies (ICMLANT), pp. 1–6 (2022). https://doi.org/10.1109/ICMLANT56191.2022.9996472
46. Liang, Y., Huang, R., Li, X., Yang, J., Zhang, X., Wang, X.: Adaptive federated learning for electric power inspection with UAV system. In: Proceedings of the 2022 3rd International Conference on Computer Engineering and Intelligent Control Virtual Event, Chongqing, China, November 25-27, 2022., Virtual Event, Chongqing, China, November 25-27, 2022. CEUR Workshop Proceedings, vol. 3344, pp. 234–241 (2022). http://ceur-ws.org/Vol-3344/paper33.pdf
47. Bian, J., Shen, C., Xu, J.: Joint Client Assignment and UAV Route Planning for Indirect-Communication Federated Learning. arXiv. arXiv:2304.10744 [cs] (2023). https://doi.org/10.48550/arXiv.2304.10744
48. Huang, J.: Maximum likelihood estimation of dirichlet distribution parameters. CMU Technique report **18** (2005)
49. 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
50. Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S., Schiele, B.: The cityscapes dataset for semantic urban scene understanding. CoRR, 3213–3223 (2016) arXiv:1604.01685
51. Sharafaldin, I., Lashkari, A.H., Ghorbani, A.A.: Toward generating a new intrusion detection dataset and intrusion traffic characterization. In: Proceedings of the 4th International Conference on Information Systems Security and Privacy (ICISSP 2018), pp. 108–116. SciTePress, Lisbon, Portugal (2018). https://doi.org/10.5220/0006639801080116
52. Kyrkou, C., Theocharides, T.: Deep-learning-based aerial image classification for emergency response applications using unmanned aerial vehicles. CoRR (2019) arXiv:1906.08716
53. Assefa, T., Zhang, J., Chowda-Reddy, R.V., Moran Lauter, A.N., Singh, A., O'Rourke, J.A., Graham, M.A., Singh, A.K.: Deconstructing the genetic architecture of iron deficiency chlorosis in soybean using genome-wide approaches. BMC Plant Biology **20**(42) (2020) https://doi.org/10.1186/s12870-020-2237-5
54. Shamsoshoara, A., Afghah, F., Razi, A., Zheng, L., Fulé, P.Z., Blasch, E.: Aerial imagery pile burn detection using deep learning: the FLAME dataset. CoRR (2020) arXiv:2012.14036
55. Xia, G.-S., Hu, J., Hu, F., Shi, B., Bai, X., Zhong, Y., Zhang, L., Lu, X.: Aid: A benchmark data set for performance evaluation of aerial scene classification. IEEE Trans. Geosci. Remote Sens. **55**(7), 3965–3981 (2017). https://doi.org/10.1109/tgrs.2017.2685945
56. Wang, B., Zhang, L., Wen, L., Liu, X., Wu, Y.: Towards real-world prohibited item detection: A large-scale x-ray benchmark. In: IEEE International Conference on Computer Vision (ICCV) (2021). https://doi.org/10.1109/ICCV48922.2021.00536

57. Zou, Q., Ni, L., Zhang, T., Wang, Q.: Deep learning based feature selection for remote sensing scene classification. IEEE Geosci. Remote Sens. Lett. **12**(11), 2321–2325 (2015). https://doi.org/10.1109/LGRS.2015.2475299
58. Netzer, Y., Wang, T., Coates, A., Bissacco, A., Wu, B., Ng, A.Y.: Reading digits in natural images with unsupervised feature learning. In: NIPS Workshop on Deep Learning and Unsupervised Feature Learning 2011 (2011). http://ufldl.stanford.edu/housenumbers/nips2011_housenumbers.pdf
59. Krizhevsky, A., Hinton, G., et al.: Learning multiple layers of features from tiny images. Master's thesis, Department of Computer Science, University of Toronto (2009)
60. Xiao, H., Rasul, K., Vollgraf, R.: Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms. CoRR (2017). https://doi.org/10.48550/arXiv.1708.07747
61. Caldas, S., Wu, P., Li, T., Konečný, J., McMahan, H.B., Smith, V., Talwalkar, A.: LEAF: A benchmark for federated settings. CoRR (2018) arXiv:1812.01097
62. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proc. IEEE **86**(11), 2278–2324 (1998). https://doi.org/10.1109/5.726791
63. Vepakomma, P., Gupta, O., Swedish, T., Raskar, R.: Split learning for health: Distributed deep learning without sharing raw patient data. CoRR (2018). https://doi.org/10.48550/arXiv.1812.00564
64. Qi, P., Chiaro, D., Guzzo, A., Ianni, M., Fortino, G., Piccialli, F.: Model aggregation techniques in federated learning: A comprehensive survey. Future Gener. Comput. Syst. **150**, 272–293 (2024). https://doi.org/10.1016/j.future.2023.09.008
65. Grasso, C., Raftopoulos, R., Schembra, G., Serrano, S.: H-HOME: A learning framework of federated FANETs to provide edge computing to future delay-constrained IoT systems. Comput. Netw. **219**, 109449 (2022). https://doi.org/10.1016/j.comnet.2022.109449
66. Jin, Z., Zhou, J., Li, B., Wu, X., Duan, C.: Fl-iids: A novel federated learning-based incremental intrusion detection system. Future Gener. Comput. Syst. **151**, 57–70 (2024). https://doi.org/10.1016/j.future.2023.09.019
67. Wuerkaixi, A., Cui, S., Zhang, J., Yan, K., Han, B., Niu, G., Fang, L., Zhang, C., Sugiyama, M.: Accurate forgetting for heterogeneous federated continual learning. In: The Twelfth International Conference on Learning Representations (2024). https://openreview.net/forum?id=ShQrnAsbPI
68. Bhuyan, N., Moharir, S.: Multi-model federated learning. In: 2022 14th International Conference on COMmunication Systems & NETworkS (COMSNETS), pp. 779–783 (2022). https://doi.org/10.1109/comsnets53615.2022.9668435. IEEE