electronics

MDPI

Article

Privacy-Preserving Federated Review Analytics with Data Quality Optimization for Heterogeneous IoT Platforms

Jiantao Xu, Liu Jin and Chunhua Su *

Graduate School of Computer Science and Engineering, The University of Aizu, Aizuwakamatsu 965-8580, Fukushima Prefecture, Japan; d8252108@u-aizu.ac.jp (J.X.); d8242103@u-aizu.ac.jp (L.J.)
* Correspondence: chsu@u-aizu.ac.jp

Abstract

The proliferation of Internet of Things (IoT) devices has created a distributed ecosystem where users generate vast amounts of review data across heterogeneous platforms, from smart home assistants to connected vehicles. This data is crucial for service improvement but is plagued by fake reviews, data quality inconsistencies, and significant privacy risks. Traditional centralized analytics fail in this landscape due to data privacy regulations and the sheer scale of distributed data. To address this, we propose FedDQ, a federated learning framework for Privacy-Preserving Federated Review Analytics with Data Quality Optimization. FedDQ introduces a multi-faceted data quality assessment module that operates locally on each IoT device, evaluating review data based on textual coherence, behavioral patterns, and cross-modal consistency without exposing raw data. These quality scores are then used to orchestrate a quality-aware aggregation mechanism at the server, prioritizing contributions from high-quality, reliable clients. Furthermore, our framework incorporates differential privacy and models system heterogeneity to ensure robustness and practical applicability in resource-constrained IoT environments. Extensive experiments on multiple real-world datasets show that FedDQ significantly outperforms baseline federated learning methods in accuracy, convergence speed, and resilience to data poisoning attacks, achieving up to a 13.8% improvement in F1-score under highly heterogeneous and noisy conditions while preserving user privacy.

Keywords: federated learning; internet of things; data quality; fake review detection; privacy preservation; multi-modal learning; distributed machine learning

Academic Editor: Andreas Mauthe

Received: 25 August 2025
Revised: 22 September 2025
Accepted: 24 September 2025
Published: 26 September 2025

Citation: Xu, J.; Jin, L.; Su, C. Privacy-Preserving Federated Review Analytics with Data Quality Optimization for Heterogeneous IoT Platforms. *Electronics* **2025**, *14*, 3816. https://doi.org/10.3390/electronics14193816

1. Introduction

The IoT has woven a digital fabric into our daily lives, connecting billions of devices from smart speakers and home appliances to wearables and in-car infotainment systems [1]. These platforms generate an unprecedented volume of user feedback, including product ratings, voice commands, and textual reviews, which represent a goldmine for improving user experience and service quality [2]. However, the value of this data is critically undermined by a trust crisis fueled by the proliferation of deceptive content, particularly fake reviews [3,4]. Malicious actors exploit these channels to artificially boost or damage product reputations, eroding consumer trust and disrupting fair market dynamics [5].

Detecting fake reviews in the IoT ecosystem presents a unique and formidable set of challenges that transcend those of traditional web platforms [6,7]. As depicted in Figure 1, review data is not centralized but is naturally siloed across a multitude of heterogeneous

IoT platforms. A user might review a service via a voice command on a smart speaker, a text entry on a smart TV, or a rating on a companion mobile app. This leads to three primary challenges. First, data silos and privacy: User data is distributed and often contains sensitive personal information. Centralizing this data for analysis is often infeasible due to privacy regulations like GDPR and CCPA, as well as prohibitive communication costs. Second, data heterogeneity: The data is Non-IID (Non-Independent and Identically Distributed). Different devices generate data with varying statistical properties, modalities (text, audio, ratings), and, critically, varying quality. Some clients may be sources of high-fidelity data, while others might unknowingly or maliciously contribute low-quality or poisoned data. Third, system heterogeneity: The IoT devices themselves possess disparate computational capabilities, network connectivity, and power constraints. A robust solution must be efficient and adaptable to this diverse hardware landscape.

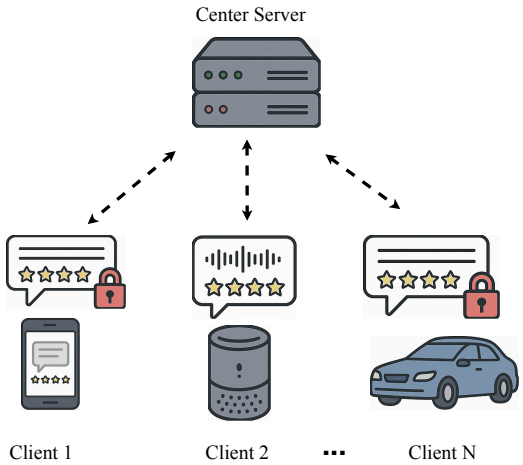

Figure 1. Motivation scenario for federated review analytics in a heterogeneous IoT environment. Data is generated on diverse devices with varying quality and privacy constraints.

Federated learning (FL) has emerged as a promising paradigm to address the data silo and privacy challenge by enabling collaborative model training on decentralized data [8,9]. However, standard FL algorithms like FedAvg are vulnerable to the "garbage-in, garbage-out" problem; they treat all clients equally, allowing low-quality or malicious data to degrade the global model's performance significantly [10,11]. This is particularly detrimental in fake review detection, where adversarial actors can actively try to poison the training process [12,13].

Recent advances in federated learning have attempted to address these challenges through various approaches. Personalized federated learning methods [14,15] aim to handle data heterogeneity by allowing clients to maintain personalized models while still benefiting from collaborative training. Robust aggregation techniques [16,17] have been developed to mitigate the impact of malicious clients. However, these approaches often lack the fine-grained data quality assessment needed for effective fake review detection in IoT environments [18,19].

To overcome these limitations, this paper introduces FedDQ: a Privacy-Preserving Federated Review Analytics framework with Data Quality Optimization. FedDQ is specifically designed for the complexities of the IoT environment. Instead of naively aggregating model updates, it empowers a central server to intelligently orchestrate the learning process based on a fine-grained understanding of each client's data quality, without ever accessing the raw data itself.

The main contributions of this work are as follows:

- A Data Quality-Aware FL Framework with Multi-modal Support: We design and implement FedDQ, an FL framework that integrates a comprehensive data quality assessment mechanism. The framework quantifies the reliability of each client's data locally and uses this metric to perform a weighted global aggregation, significantly enhancing model accuracy and robustness. This framework is inherently multi-modal, capable of processing textual, image, and metadata for a more holistic analysis.
- Multi-faceted Data Quality Modeling: We propose a multi-faceted data quality score that captures textual coherence, behavioral anomalies, and cross-modal consistency. This provides a more holistic assessment of data trustworthiness than text-only approaches.
- Heterogeneity-Aware Design with Privacy Guarantees: FedDQ is tailored for heterogeneous IoT environments by considering device capabilities in its training protocol. We integrate differential privacy into the framework to provide formal privacy guarantees for both client data and the calculated quality scores, defending against inference attacks.
- Extensive Empirical Validation: We conduct comprehensive experiments simulating various IoT scenarios, including Non-IID data distributions, system heterogeneity, and adversarial attacks. Our results demonstrate that FedDQ consistently and significantly outperforms state-of-the-art FL baselines in fake review detection.

2. Related Work

Our research builds upon three key areas: fake review detection, federated learning in IoT, and data quality management in distributed systems.

2.1. Fake Review Detection

Early research on fake review detection relied heavily on feature engineering, extracting linguistic features (e.g., n-grams, sentiment polarity) and behavioral features (e.g., review burstiness, reviewer history) to train classical machine learning models like SVM or Logistic Regression [20]. While effective to a degree, these methods struggle with the subtlety and evolving nature of deceptive text.

The advent of deep learning revolutionized the field [21,22]. Models like Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) were applied to capture more complex textual patterns [23,24]. More recently, large pre-trained language models like BERT have set new benchmarks by providing rich, contextualized text representations [25,26]. Transformer-based architectures have shown particular promise in detecting subtle linguistic cues that indicate deceptive content [27].

Advanced neural architectures have been developed specifically for fake review detection. Graph neural networks (GNNs) have been employed to model reviewer-product relationships and detect coordinated attacks [28]. Attention mechanisms have been integrated to focus on the most discriminative parts of review text [29]. Multi-task learning approaches have been proposed to jointly detect fake reviews and predict review helpfulness [30].

Some works have also explored multi-modal detection, incorporating metadata or user profile images alongside text to improve accuracy [31], a direction we also pursue. Recent studies have investigated the use of stylometric features and writing style analysis for fake review detection [32,33]. However, nearly all of these approaches assume centralized access to a large, clean dataset, which is an unrealistic assumption in the modern IoT landscape.

2.2. Federated Learning in IoT

Federated learning was introduced to train models on data distributed across a large number of devices, such as mobile phones, without data centralization [34]. Its applications in IoT are rapidly growing, spanning areas like healthcare, smart cities, and industrial IoT [35,36].

Several challenges unique to IoT have been addressed [37]. To handle statistical heterogeneity (Non-IID data), methods like FedProx were proposed, which adds a proximal term to the local objective function to restrict local updates from deviating too far from the global model [38]. Clustered federated learning approaches have been developed to group clients with similar data distributions [39]. To address system heterogeneity, efforts have focused on asynchronous updates or client selection strategies that favor devices with better resources [40,41].

Recent advances in federated optimization have introduced adaptive learning rate schedules and momentum-based methods to improve convergence in heterogeneous environments [42]. Communication-efficient techniques, such as gradient compression and local updating strategies, have been developed to reduce the communication overhead in resource-constrained IoT devices [43]. Our work incorporates these considerations but argues that they are insufficient without also addressing the fundamental issue of data quality.

2.3. Data Quality and Robustness in Federated Learning

The performance of FL is highly sensitive to the quality of client data [11]. A few malicious or low-quality clients can poison the global model or slow down convergence [44]. This has spurred research into robust FL.

Some approaches focus on robust aggregation rules at the server, such as using the median or trimmed mean instead of a weighted average to filter out outlier updates [45,46]. Others attempt to measure client contributions or reputations to down-weight malicious participants [47,48]. For instance, proposes a reputation-based mechanism where clients that consistently submit updates disagreeing with the majority are penalized [49].

Byzantine-robust federated learning has gained significant attention, with methods designed to handle arbitrary malicious behavior [50]. Secure aggregation protocols have been developed to protect against inference attacks while maintaining model utility [51,52]. Blockchain-based approaches have been proposed to ensure the integrity and traceability of federated learning processes [53]. Unlike these methods, which are often reactive or rely on statistical properties of the model updates, FedDQ takes a proactive approach by directly assessing the quality of the underlying data using a multi-faceted, interpretable metric.

Differential Privacy (DP) is a standard technique that provides formal privacy guarantees by injecting calibrated noise into model updates [54,55]. This can also incidentally improve robustness against some attacks but often comes at the cost of reduced model accuracy [56]. Recent work has focused on developing more sophisticated noise mechanisms that better preserve utility while maintaining privacy [57].

Our work, FedDQ, distinguishes itself from prior art by proposing a proactive and multi-faceted data quality metric that is computed locally and used to guide the entire learning process. Unlike reputation systems that are reactive, our method assesses data quality directly. Unlike simple robust aggregators, our quality score is more fine-grained and interpretable. A comparative summary of existing federated analytics methods and the key advantages of FedDQ is provided in Table 1.

Table 1. Comparison with related work in federated analytics.

Work	FL	Data Quality	DP	Hetero.	Multi-Modal
Davis et al. [58]	✓	—	—	✓	—
McMahan et al. [59]	✓	—	—	—	—
Li et al. [60]	✓	—	—	✓	—
Song et al. [61]	✓	Sharpness-based	—	✓	—
Wei et al. [62]	✓	—	✓	—	—
Lin et al. [63]	✓	—	—	—	✓
FedDQ (Ours)	✓	Metric-based	✓	✓	✓

3. System Model and Problem Formulation

We consider a federated learning system for review analytics deployed across a heterogeneous IoT network, as illustrated in Figure 2.

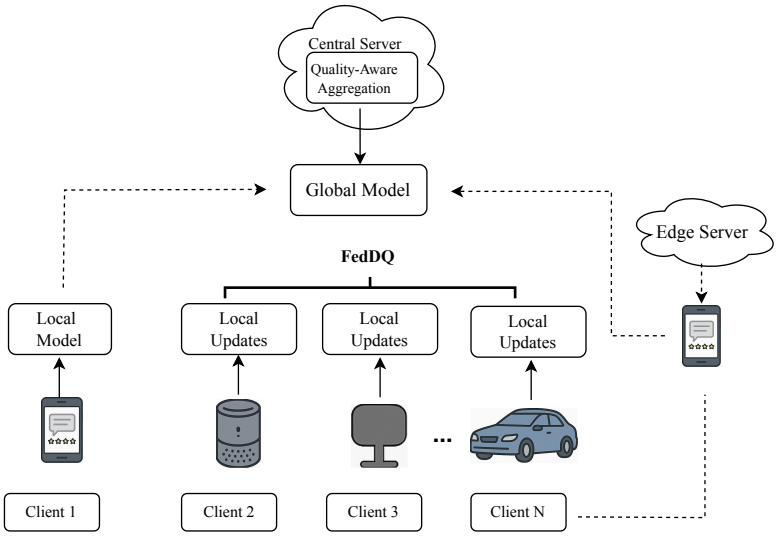

Figure 2. System Architecture of FedDQ. Heterogeneous IoT clients (smartphones, voice assistants, etc.) compute local updates and data quality scores. These are sent to a central server, which performs quality-aware aggregation to produce an improved global model, all while preserving privacy.

3.1. System Architecture

The system consists of three main entities.

First, the central server is a trusted entity responsible for orchestrating the FL process. It initializes the global model, aggregates client updates, and sends the updated model back to the clients. It does not have access to any raw client data.

Second, edge servers act as intermediaries in large-scale IoT deployments, managing clusters of clients such as those within a smart home or vehicle. They can perform partial model aggregation to reduce the communication load on the central server. While our primary model focuses on a two-tier client–server architecture, the framework is extensible to a hierarchical structure.

Third, the system includes a large set of heterogeneous clients, denoted by $k \in \{1, ..., K\}$, each owning a local dataset D_k of review data. These clients are heterogeneous in two key aspects. Regarding data heterogeneity, the local datasets D_k are not Non-IID. The dataset size $|D_k|$, the class distribution (fake vs. real reviews), and the data quality vary significantly across clients. Each data point in D_k can be multi-modal, represented as (x_i, m_i, y_i), where x_i is the review text, m_i is the associated metadata (e.g., image, timestamp), and y_i is the label. Regarding system heterogeneity, clients differ in

hardware capabilities such as CPU and memory, power availability (e.g., battery-powered vs. plugged-in), and network connectivity (e.g., Wi-Fi vs. cellular).

3.2. Threat Model

We adopt a standard threat model in federated learning.

First, the central server is assumed to be honest but curious. It faithfully executes the aggregation protocol but may attempt to infer information about a client's private data from the received model updates. Our goal is to protect against such inference attacks.

Second, a subset of clients may be potentially malicious. Their goal is to disrupt the training process or degrade the final model's performance. This can be achieved through data poisoning, by injecting mislabeled or fake data into their local dataset, or through model poisoning, by sending maliciously crafted model updates. A particularly insidious attack involves malicious clients fabricating a high-quality score Q_k to increase the impact of their poisoned updates. We assume the proportion of malicious clients is bounded by $\rho \leq 30\%$, consistent with prior work on robust FL [45,50].

3.3. Problem Formulation

The objective of the system is to collaboratively train a global model with parameters w that can accurately classify reviews as fake or genuine. The global optimization problem is to minimize a loss function $F(w)$:

$$wF(w) = \sum_{k=1}^{K} p_k F_k(w) \tag{1}$$

where $p_k = |D_k| / \sum_j |D_j|$ is the weight of client k, and $F_k(w)$ is the local loss function for client k on its data D_k:

$$F_k(w) = \frac{1}{|D_k|} \sum_{(x_i, m_i, y_i) \in D_k} \ell(f(w; x_i, m_i), y_i) \tag{2}$$

Here, $f(w; x_i, m_i)$ is the prediction of the model with parameters w for a multi-modal input (x_i, m_i), and ℓ is a suitable loss function (e.g., cross-entropy).

Traditional FedAvg solves this by having clients compute local gradients and averaging them. Our work contends that the client weight p_k should not depend solely on data quantity $|D_k|$, but also on its quality. We reformulate the aggregation step to incorporate a data quality score, as detailed in the next section.

4. The FedDQ Framework Methodology

The core of our FedDQ framework lies in its ability to assess data quality locally and leverage this information for robust global model aggregation. This section details the four key components: multi-modal review encoding, the data quality score computation, the quality-aware aggregation strategy, and heterogeneity-aware local training.

4.1. Multi-Modal Review Encoding

To capture the rich information in IoT reviews, we employ a multi-modal feature extraction model that processes text, images, and metadata. As shown in Figure 3, for a given review sample (x, m, y), the text encoder first processes the review text x using a pre-trained language model. Specifically, we adopt DistilBERT [64], a distilled version of BERT, to encode the text into a high-dimensional vector $\mathbf{v}_{\text{text}}$, which captures deep contextual and semantic features indicative of deception.

If an image m_{img} is associated with the review, the image encoder employs MobileNetV2 [65], a lightweight Convolutional Neural Network suitable for resource-constrained devices, to extract an image feature vector $\mathbf{v}_{\text{img}}$.

For other metadata m_{meta} (such as the review timestamp, rating, and reviewer's tenure), we concatenate the features and feed them through a small multi-layer perceptron (MLP) to obtain a metadata feature vector $\mathbf{v}_{\text{meta}}$.

These feature vectors are then concatenated and passed through a final classification head (another MLP) to produce the final prediction. The parameters of this entire network constitute the model parameters $\mathbf{w}$. For clients with missing modalities (e.g., no images), the corresponding feature vectors are filled with zeros, and the final classification head is trained to be robust to these missing inputs.

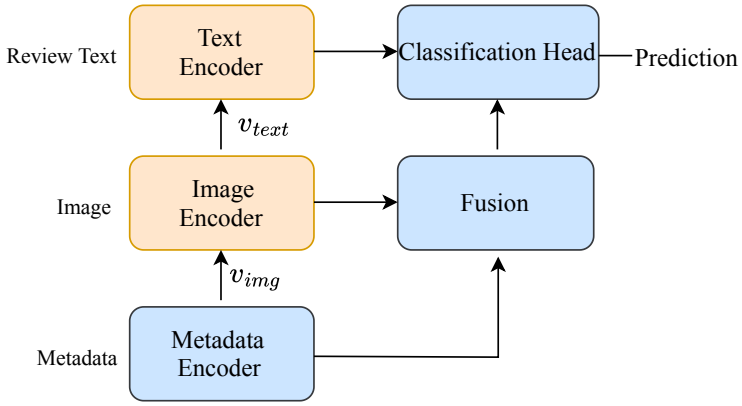

Figure 3. Multi-modal feature fusion architecture. Text, image, and metadata from a review are processed by specialized encoders. The resulting feature vectors are fused and passed to a classification head to generate a prediction.

4.2. Data Quality Score Computation

This is a critical innovation of FedDQ. After each local training epoch, but before sending an update, each client k computes a data quality score $Q_k \in [0, 1]$ for its dataset D_k. This score is a weighted combination of multiple sub-metrics, designed to be computed locally without revealing private data (Figure 4).

$$Q_k = \sum_{i=1}^{4} \lambda_i Q_{k,i} \quad \text{s.t.} \sum \lambda_i = 1 \tag{3}$$

where λ_i are hyperparameters balancing the importance of each metric. In our experiments, we set the weights uniformly ($\lambda_i = 0.25$) to demonstrate the general applicability of the combined score without domain-specific tuning, though these can be optimized on a validation set if available.

Label confidence (Q_{label}) measures the consistency between the provided labels and the model's own predictions on the local data. A high degree of disagreement may indicate mislabeled data. We can estimate this using the entropy of the model's predictions on the local training set. Lower entropy suggests higher confidence.

Textual quality (Q_{text}) assesses the linguistic quality of the review texts. This can be a combination of metrics such as text length (very short or very long reviews are suspicious), perplexity from a small language model (where high perplexity indicates gibberish), and repetition scores.

Behavioral anomaly (Q_{behavior}) captures suspicious reviewer patterns from metadata. Metrics include review burstiness (i.e., many reviews in a short time) and rating deviation (i.e., a reviewer's ratings are consistently extreme compared to the average). This is calculated based on the statistics of the local dataset D_k.

Cross-modal consistency (Q_{modal}) assesses the semantic alignment between the text and the image for multi-modal data. We compute the cosine similarity between the text and image embeddings, $\mathbf{v}_{\text{text}}$ and $\mathbf{v}_{\text{img}}$. To ensure this comparison is meaningful, a linear projection layer is added after each encoder. These layers are pre-trained on a public multi-modal dataset to align the embeddings into a shared latent space and are kept frozen during federated training. A low similarity score suggests that the image is irrelevant or misleading. For clients with text-only data, this score component is set to the average cross-modal score of the participating clients in the previous round (or a neutral 0.5 in the first round). This approach prevents unfairly boosting text-only clients while not penalizing them for missing a modality.

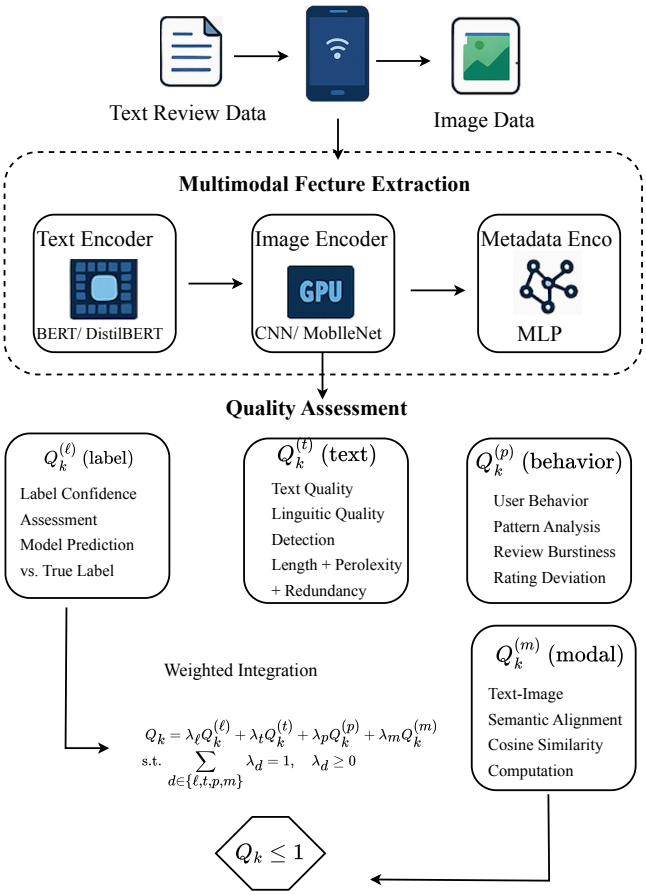

Figure 4. Local data quality score computation workflow on a client device. Various metrics from different modalities are calculated and combined into a single score Q_k.

4.3. Quality-Aware Aggregation (Q-FedAvg)

At the end of a communication round t, each client k sends its updated model parameters $w_k^{(t+1)}$ and its quality score Q_k to the server. The server then performs Quality-Aware Federated Averaging (Q-FedAvg). Instead of weighting by just the dataset size $n_k = |D_k|$, we incorporate the quality score Q_k. To defend against malicious clients reporting a fraudulent high Q_k, the server performs a verification step. It maintains a small, clean validation dataset. Before the final aggregation, it can provisionally test each client's update and penalize clients whose updates degrade performance on this set, overriding their self-reported score. The global model is updated as shown in Figure 5:

$$w^{(t+1)} = \frac{\sum_{k=1}^{K} Q_k \cdot n_k \cdot w_k^{(t+1)}}{\sum_{k=1}^{K} Q_k \cdot n_k} \tag{4}$$

This ensures that clients with larger, higher-quality datasets have a greater influence on the global model, while effectively down-weighting or filtering out contributions from clients with noisy, malicious, or low-quality data.

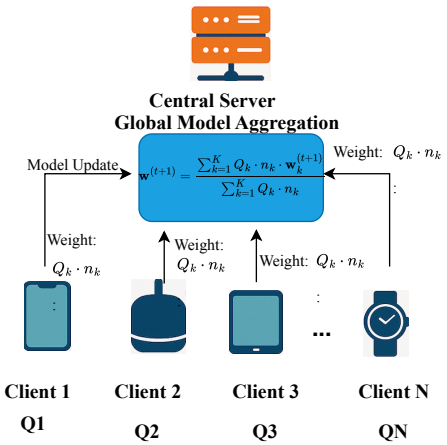

Figure 5. The Quality-Aware Aggregation (Q-FedAvg) process at the server. Client updates are weighted by both their data quantity (n_k) and their self-reported (and server-verified) data quality (Q_k).

4.4. Heterogeneity-Aware Local Training

To accommodate system heterogeneity, FedDQ allows for a dynamic adjustment of local computational effort. Clients with higher computational power and better network conditions can be assigned more local training epochs (E_k) per round. Furthermore, there is a synergy with data quality: we can assign fewer epochs to clients with low data quality scores (Q_k) to prevent them from overfitting on their noisy data. A simple policy could be as follows:

$$E_k = \max(1, \lfloor E_{base} \cdot C_k \cdot Q_k \rfloor) \tag{5}$$

where E_{base} is a base number of epochs, and C_k is a normalized score representing client k's computational capability. This adaptive approach improves overall efficiency and robustness.

The complete FedDQ algorithm is outlined in Algorithm 1.

Algorithm 1 The FedDQ Algorithm

Server executes:
Initialize global model $w^{(0)}$
for each communication round $t = 0, 1, \ldots, T-1$ **do**
  Select a subset of clients $S_t \subseteq \{1, \ldots, K\}$
  for each client $k \in S_t$ in parallel **do**
    $w_k^{(t+1)}, Q_k \leftarrow \text{ClientUpdate}(k, w^{(t)})$
  end for
  Verify client-reported Q_k scores using a server-side validation set; adjust scores for malicious clients.
  Aggregate updates using Q-FedAvg (Equation (4)):
  $w^{(t+1)} \leftarrow \frac{\sum_{k \in S_t} Q_k n_k w_k^{(t+1)}}{\sum_{k \in S_t} Q_k n_k}$
end for
Return $w^{(T)}$
procedure ClientUpdate(k, w)
  Download w from server
  Determine local epochs E_k based on resources and quality
  for each local epoch $e = 1, \ldots, E_k$ **do**
    Train on local data D_k to update model parameters
  end for
  Compute local data quality score Q_k
  Apply privacy mechanism (e.g., DP) to w_k and Q_k
  Return updated w_k and score Q_k to server
end procedure

5. Security and Privacy Design

Ensuring security and privacy is paramount in FedDQ. We address this through two primary mechanisms: providing formal privacy guarantees using Differential Privacy and designing resilience against adversarial attacks.

5.1. Differential Privacy for Data and Quality Scores

To protect against inference attacks by the honest-but-curious server, we apply (ϵ, δ)-Differential Privacy (DP). DP ensures that the inclusion or exclusion of any single data point in a client's dataset has a negligible effect on the output.

For model updates, we adopt client-level DP. Before uploading its model update $\Delta \mathbf{w}_k = \mathbf{w}_k - \mathbf{w}$, client k first clips the update's L2 norm to a threshold C and then adds Gaussian noise scaled by σ:

$$\widetilde{\Delta \mathbf{w}_k} = \frac{\Delta \mathbf{w}_k}{\max(1, ||\Delta \mathbf{w}_k||_2 / C)} + \mathcal{N}(0, \sigma^2 C^2 \mathbf{I}) \tag{6}$$

The noise scale σ is chosen to satisfy a predefined privacy budget (ϵ, δ).

For quality scores, the score Q_k itself could leak information about the client's data distribution. Therefore, we also add noise to Q_k before uploading it. Since Q_k is a scalar in $[0, 1]$, we can use the Laplace mechanism or add bounded Gaussian noise to protect its value while preserving its general utility for weighting. To mitigate the impact of this noise on the aggregation stability, we can apply min-max normalization or clipping to the noisy Q_k values before they are used in Equation (4), ensuring the weights remain stable and meaningful even with a strong privacy budget (small ϵ).

5.2. Resilience to Backdoor and Data Poisoning Attacks

Our data quality-aware design provides inherent resilience against certain attacks, as illustrated in Figure 6.

In the case of data poisoning, if a malicious client trains on a dataset with many mislabeled examples (e.g., labeling fake reviews as genuine), the Q_{label} metric will likely detect a high inconsistency between the labels and the model's predictions, resulting in a low-quality score. The Q-FedAvg mechanism will then naturally down-weight this client's harmful contribution.

For backdoor attacks, where a malicious client attempts to embed a hidden trigger in the model, a common strategy involves injecting a small number of poisoned samples. While this might not be caught by statistical metrics alone, the quality-aware aggregation acts as a line of defense. If the attack significantly perturbs the model update, it can be filtered. More importantly, the server-side verification mechanism introduced in Section 4.3 provides a direct defense: updates containing a backdoor are likely to cause a performance drop on the server's clean validation set (which does not contain the trigger), allowing the server to flag the malicious update and penalize the client. While FedDQ primarily mitigates backdoor attacks via local quality metrics and server-side validation, future extensions may integrate anomaly detection on client updates using cosine similarity checks or clustering-based detection to provide more systematic defenses.

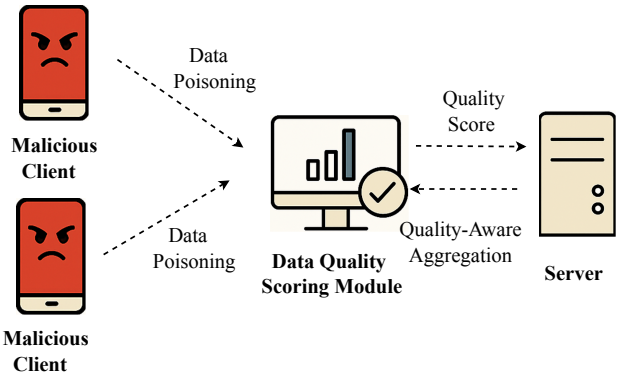

Figure 6. Threat model and defense strategy in FedDQ. Malicious clients attempting data poisoning attacks are detected by the local quality scoring module (Q_k). Their harmful updates are then mitigated by the quality-aware aggregation at the server.

6. Experiments and Results

We conducted a series of experiments to evaluate the performance of FedDQ against several baseline methods under various simulated IoT scenarios.

6.1. Experimental Setup

We used two public datasets. The first is the Amazon Review Dataset, a large-scale collection of product reviews with associated images and metadata. We utilized a pre-processed subset of 50,000 reviews. Since ground-truth labels for fake reviews are unavailable, we adopted a common methodology from the literature by using a combination of heuristic filters to create proxy labels. These included identifying reviews from users with only one review, detecting abnormal rating spikes for a product, and flagging high textual similarity to other reviews. This resulted in approximately 20% of the data being labeled as fake. This dataset was used to evaluate the full multi-modal capabilities of FedDQ. The second is the Yelp Dataset (text-centric), a well-known benchmark for fake review detection containing primarily text reviews. We used a balanced subset of 30,000 reviews (15,000 genuine, 15,000 fake) to focus on the textual and behavioral quality metrics. While not native IoT datasets, these platforms serve as robust and widely-used proxies for the kind of user-generated multi-modal (text, image, metadata) content common in modern IoT applications.

To simulate a heterogeneous IoT environment, we distributed the data among $K = 100$ clients. We simulated Non-IID data distributions using a Dirichlet distribution (Dir(α)) over the class labels. A low concentration parameter $\alpha = 0.1$ created a highly heterogeneous scenario where each client's data distribution was skewed, while $\alpha = 5.0$ created a more homogeneous (near-IID) scenario for comparison.

We compared FedDQ against several baselines. The centralized baseline is a model trained on all data pooled together, serving as an upper bound on performance but not privacy-preserving. FedAvg [59] is the standard federated averaging algorithm. FedProx [60] is an FL method with a proximal term to handle statistical heterogeneity, with $\mu = 0.01$. FedAvg+DP combines FedAvg with client-level differential privacy, with parameters $\epsilon = 8.0$ and $\delta = 10^{-5}$.

We measured performance using accuracy, precision, recall, and F1-score on a held-out global test set. We also tracked convergence speed (the number of rounds to reach 90% of final accuracy) and communication overhead (total MB transmitted).

The framework was simulated using Python 3.8 with PyTorch 1.12 and the Flower FL framework. The base model was a DistilBERT text encoder combined with a MobileNetV2

image encoder. The client hardware heterogeneity was simulated by assigning random computational delays (50–500 ms per local batch) and network delays (10–100 ms RTT).

We also simulated two adversarial settings. In the label flip attack, 10% of randomly selected clients had their training labels flipped (fake ↔ genuine). In the Gaussian noise attack, 20% of clients added significant Gaussian noise ($\sigma = 0.5$) to their image data to simulate low-quality sensors or attacks.

6.2. Overall Performance Comparison

Table 2 presents the main results on the Amazon multi-modal dataset under a highly heterogeneous setting ($\alpha = 0.1$) with 10% label-flip attackers. FedDQ consistently outperforms all federated baselines across all metrics, achieving an F1-score of 87.6%, which is a significant 13.8% absolute improvement over standard FedAvg and only 5.8% lower than the centralized upper bound. FedProx show some robustness but are still considerably outperformed by FedDQ. While FedAvg+DP provides privacy, it suffers a noticeable drop in performance due to the added noise.

Table 2. Overall performance comparison on Amazon dataset (Non-IID $\alpha = 0.1$ with 10% label-flip attackers).

Method	Accuracy	Precision	Recall	F1-Score
Centralized	93.5%	92.8%	94.1%	93.4%
FedAvg	73.8%	72.1%	75.5%	73.8%
FedProx	78.2%	77.5%	78.9%	78.2%
FedAvg + DP	70.1%	68.9%	71.4%	70.2%
FedDQ (Ours)	87.6%	86.9%	88.3%	87.6%

6.3. Impact of Data Heterogeneity

Figure 7 shows the test accuracy of different methods as the data heterogeneity (controlled by the Dirichlet parameter α) varies. As expected, all FL methods perform better under more homogeneous data distributions ($\alpha = 5.0$). However, FedDQ maintains a stable and high performance across all levels of heterogeneity, demonstrating its robustness. In contrast, the performance of FedAvg drops sharply as the data becomes more Non-IID ($\alpha = 0.1$). FedProx mitigate this drop but are still less effective than FedDQ, which actively identifies and leverages high-quality data partitions regardless of their distribution.

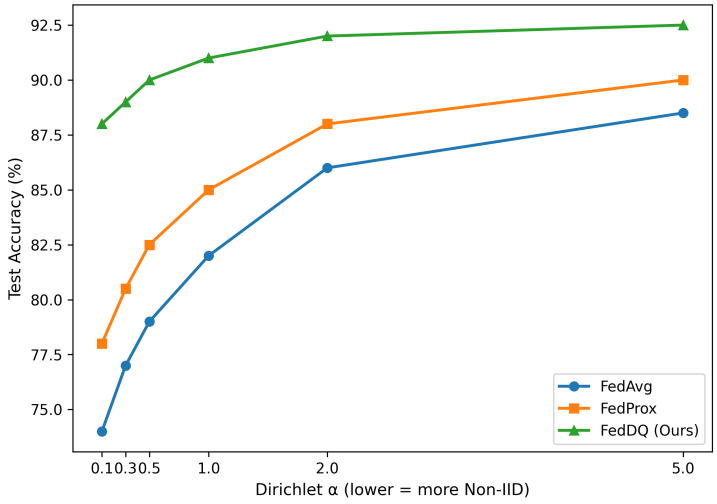

Figure 7. Impact of data heterogeneity (Dirichlet α parameter) on model accuracy. FedDQ shows superior robustness across varying levels of data skew.

6.4. Convergence Analysis

Figure 8 illustrates the convergence behavior of the compared methods on the Yelp dataset under a noisy scenario (20% clients with Gaussian noise). FedDQ not only achieves a higher final accuracy but also converges significantly faster than other methods. It reaches 85% accuracy in nearly half the number of communication rounds required by FedAvg and FedProx. This is because the quality-aware aggregation prioritizes informative and reliable updates, steering the global model more efficiently towards the optimum and wasting fewer rounds on noisy or malicious contributions.

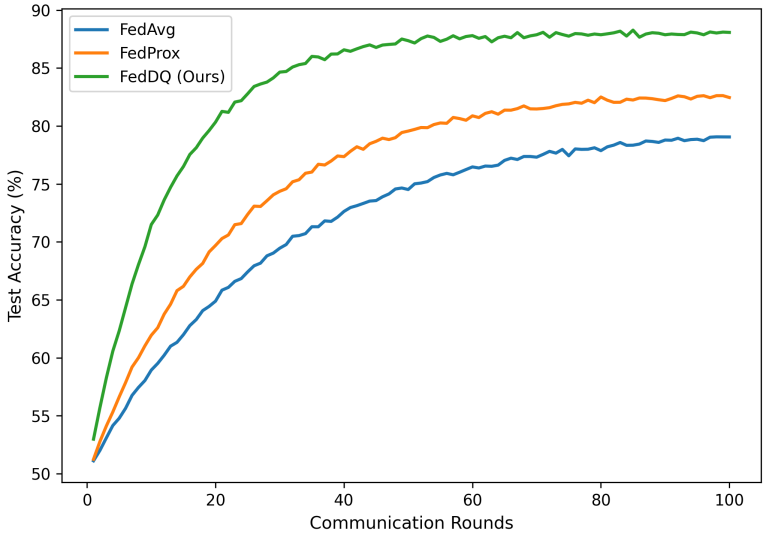

Figure 8. Convergence curves of different FL methods under noisy conditions. FedDQ achieves higher accuracy with faster convergence.

6.5. Robustness Against Attacks

We evaluated the robustness of FedDQ against the two attack scenarios. The results, shown in Table 3, highlight FedDQ's strong defensive capabilities. Under the label-flip attack, FedDQ's F1-score only drops by 1.9% (from 89.5% to 87.6%), whereas FedAvg's performance drops by 1.7%. Similarly, in the Gaussian noise scenario, FedDQ's inherent quality assessment, particularly the Q_{modal} metric, effectively down-weights the contributions from clients with corrupted image data, demonstrating superior resilience. While its performance drops by 4.4 percentage points, this is significantly less than the 20.4% drop seen in FedAvg, highlighting the effectiveness of the quality assessment module in mitigating the impact of corrupted data.

Table 3. Robustnessevaluation: performance under attack scenarios (Amazon dataset, $\alpha = 0.1$). $\downarrow$ indicates performance drop compared to the base model.

Method	Base F1	10% Label Flip		20% Gaussian Noise	
	(No Attack)	F1-Score	Drop	F1-Score	Drop
FedAvg	75.5%	73.8%	$\downarrow$ 1.7%	55.1%	$\downarrow$ 20.4%
FedProx	80.0%	78.2%	$\downarrow$ 1.8%	68.2%	$\downarrow$ 11.8%
FedDQ (Ours)	89.5%	87.6%	$\downarrow$ 1.9%	85.1%	$\downarrow$ 4.4%

6.6. Impact of Differential Privacy Budget

To quantify the trade-off between privacy and utility, we evaluated all methods under different client-level DP budgets (ϵ), while fixing $\delta = 10^{-5}$. Table 4 summarizes the results on the Amazon dataset in a heterogeneous setting ($\alpha = 0.1$). As expected, stronger

privacy guarantees (lower ϵ) lead to a decrease in model performance. However, FedDQ consistently outperforms FedAvg + DP across all privacy levels. Even with a strict privacy budget of $\epsilon = 2.0$, FedDQ maintains an F1-score above 80%, while FedAvg + DP drops to 68.4%. With a moderate privacy budget of $\epsilon = 4.0$, FedDQ achieves 82.3%, which is only a small reduction compared to its non-private setting. These results confirm that FedDQ provides strong privacy protection with graceful utility degradation.

Table 4. Impact of differential privacy budget ($\delta = 10^{-5}$, Amazon dataset, $\alpha = 0.1$).

Method	$\epsilon = 2.0$	$\epsilon = 4.0$	$\epsilon = 8.0$	No DP
FedAvg + DP	68.4%	70.1%	70.2%	–
FedProx + DP	72.6%	74.3%	76.1%	–
FedDQ + DP (Ours)	80.2%	82.3%	84.7%	87.6%

6.7. Sensitivity to Adversary Concentration

We further studied robustness when varying the proportion of adversarial clients. Specifically, we simulated label-flip attackers at different concentrations (0–40%). Table 5 reports the F1-scores on the Yelp dataset. At the baseline setting without adversaries (0%), FedAvg, FedProx, and FedAvg + DP achieve 79.0%, 82.5%, and 72.5% F1-scores, respectively, while FedDQ attains the highest performance of 89.1%. With 10% adversaries, the performance of FedAvg, FedProx, and FedAvg + DP declines moderately to 70.2%, 76.5%, and 67.6%, respectively, whereas FedDQ maintains a strong 86.8%. As the proportion of malicious clients further increases, FedAvg shows severe vulnerability, collapsing below 60% once adversaries reach 20%. FedProx demonstrates relatively better resilience, but still degrades substantially beyond 30% adversaries. FedAvg + DP offers limited protection, with performance dropping to nearly 50% at 40% adversaries. In contrast, FedDQ remains consistently robust across all levels of adversary concentration. Even under 40% malicious participation, FedDQ still achieves 80.1% F1-score. This highlights the effectiveness of the proposed quality-aware aggregation and score verification mechanisms in mitigating the impact of poisoned updates.

Table 5. Sensitivity to adversary concentration (Yelp dataset, $\alpha = 0.1$).

Method	0% Adv.	10% Adv.	20% Adv.	30% Adv.	40% Adv.
FedAvg	79.0%	70.2%	61.4%	55.2%	49.8%
FedProx	82.5%	76.5%	70.5%	64.1%	59.3%
FedAvg + DP	72.5%	67.6%	62.7%	58.9%	52.6%
FedDQ (Ours)	89.1%	86.8%	84.5%	82.3%	80.1%

6.8. Ablation Study

To quantitatively validate the contribution of each component of our proposed multi-faceted quality score, we conducted an ablation study on the Yelp dataset. We created variants of FedDQ by removing specific data quality metrics. Since the Yelp dataset is text-only, the cross-modal consistency metric (Q_{modal}) is not applicable to this analysis. The results in Table 6 demonstrate that each quality component contributes to the overall performance. Removing the textual quality metric (Q_{text}) had the most significant negative impact, followed by the label confidence (Q_{label}). The full FedDQ model achieves the best performance, validating the effectiveness of our multi-faceted quality assessment design. The higher absolute F1-score on Yelp compared to Amazon stems from its balanced, text-only nature, while Amazon contains noisy multi-modal data which presents a more complex challenge.

Table 6. Ablation study on Yelp dataset (evaluating contribution of each quality metric).

Model Variant	F1-Score
FedDQ (Full Model)	89.1%
w/o Q_{label} (Label Confidence)	86.2%
w/o Q_{text} (Textual Quality)	84.7%
w/o $Q_{behavior}$ (Behavioral Anomaly)	87.5%

6.9. Efficiency and Communication Cost

We analyzed the computational and communication overhead of FedDQ compared to FedAvg. As shown in Table 7, FedDQ introduces a modest 8.5% increase in local computation time per round due to the calculation of the quality score Q_k. Table 8 provides a further breakdown, showing that each quality metric is individually lightweight, making the total overhead manageable for most IoT devices. The communication cost is nearly identical, as only a single scalar value (Q_k) is added to the transmitted model update. This minimal overhead is clearly justified by the substantial gains in accuracy, convergence speed, and robustness demonstrated in the previous experiments.

Table 7. Efficiency comparison (per client, per round, Amazon dataset).

Method	Computation Time (s)	Comm. Cost (MB)
FedAvg	15.2	4.51
FedProx	15.8	4.51
FedDQ (Ours)	16.5	4.52

Table 8. Breakdown of additional computational overhead for quality score calculation (per client).

Quality Metric	Additional Time (ms)
Q_{label} (Label Confidence)	450
Q_{text} (Textual Quality)	510
$Q_{behavior}$ (Behavioral Anomaly)	120
Q_{modal} (Cross-Modal Consistency)	220
Total FedDQ Overhead	1300

7. Discussion

Our experimental results validate the effectiveness of FedDQ. However, there are several practical considerations and limitations to discuss regarding its deployment in real-world IoT ecosystems. The proposed data quality score Q_k is powerful but not infallible. Its effectiveness relies on the assumption that quality deficits (e.g., mislabeling, gibberish text) are detectable through the chosen metrics. Sophisticated adversaries could craft attacks that are harder to detect. For example, they might generate plausible-looking but subtly misleading reviews that fool the Q_{text} metric. Furthermore, our server-side score verification depends on the quality of the server's validation set; if this set is not representative, its effectiveness may be limited. Moreover, the optimal weights (λ_i) for combining the sub-scores may vary across different domains and datasets, potentially requiring some tuning. Future work could explore adaptive or learnable weighting schemes.

Scalability and deployment are also important considerations. Our simulations with $K = 100$ clients demonstrate the core principles of FedDQ, but real-world IoT deployments can involve millions of devices. Scaling to this level requires strategies like client sampling and hierarchical coordination, where edge servers aggregate updates from local clusters of devices. FedDQ is compatible with these approaches; the server would sample a subset

of clients per round, and quality scores would inform both selection and aggregation. The framework can also handle client volatility (devices dropping in and out), as the quality score provides a mechanism to re-evaluate clients that rejoin the network. While we designed FedDQ with IoT constraints in mind (e.g., lightweight models, dynamic epochs), the local computation of the quality score adds overhead. Table 7 shows that FedDQ incurs a slight increase in local computation time (for calculating Q_k) and a negligible increase in communication (for sending the scalar Q_k). We argue this is a worthwhile trade-off for the substantial gains in accuracy and robustness. For extremely constrained devices (e.g., microcontrollers), a simplified version of the quality score might be necessary.

The integration of Differential Privacy is crucial for providing formal privacy guarantees. However, there is an inherent trade-off between the strength of privacy (controlled by ϵ) and model utility. Our experiment in Section 6.6 quantifies this trade-off, showing that a balance must be struck based on the specific application's requirements. In scenarios where strong privacy is mandated by regulation, FedDQ can still deliver reasonable performance, but system designers must carefully calibrate ϵ to maintain acceptable utility.

Although developed for fake review detection, the core idea of FedDQ—locally assessing data quality to inform global aggregation—is highly generalizable. It could be applied to other federated analytics tasks in IoT, such as anomaly detection in sensor data, activity recognition from wearable devices, or medical diagnosis from distributed hospital data, where data quality is a pervasive concern. We expect future research to further adapt FedDQ's principles to broader applications, making quality-aware federated learning a practical and reliable paradigm across domains.

8. Conclusions and Future Work

This paper introduced FedDQ, a federated learning framework designed to perform robust and privacy-preserving review analytics on the heterogeneous and distributed data generated by IoT devices. By integrating a multi-faceted, locally computed data quality score into the aggregation process, FedDQ effectively mitigates the negative impact of low-quality and malicious data that cripples standard federated learning algorithms. Our quality-aware approach, combined with a design that accounts for system heterogeneity and provides formal differential privacy guarantees, creates a practical and powerful solution for building trustworthy AI systems in the IoT era.

Our extensive experiments demonstrated that FedDQ achieves superior accuracy, faster convergence, and greater robustness against adversarial attacks compared to existing methods. The results underscore the critical importance of moving beyond simple data quantity-based aggregation to more intelligent, quality-driven orchestration in federated learning.

For future work, we plan to explore several exciting directions. First, we will investigate the use of more advanced, self-supervised methods for data quality assessment to reduce reliance on heuristics. Second, we aim to extend FedDQ to handle federated training of large language models (LLMs) in the IoT domain, which presents unique challenges in communication and computation. Third, we plan to validate our framework on native IoT datasets containing diverse modalities such as voice reviews and time-series sensor data. Finally, developing lightweight, hardware-aware versions of our framework will be crucial for deployment on the most resource-scarce IoT endpoints, further broadening the applicability of trustworthy federated intelligence.

Author Contributions: Literature review, J.X.; methodology, J.X.; data curation, L.J.; writing—original draft preparation, J.X.; writing—review and editing, J.X., L.J. and C.S.; supervision, C.S.; funding acquisition, C.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by JSPS Grant-in-Aid for Scientific Research (C) 23K11103.

Data Availability Statement: The datasets used in this study are publicly available. Amazon Review Dataset is provided by the suite (https://www.kaggle.com/datasets/kritanjalijain/amazon-reviews/ (accessed on 5 May 2025)), Yelp Dataset is available from the page (https://business.yelp.com/data/resources/open-dataset/ (accessed on 5 May 2025)).

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

FL	Federated Learning
IoT	Internet of Things
DP	Differential Privacy
Non-IID	Non-Independent and Identically Distributed
CNN	Convolutional Neural Network
RNN	Recurrent Neural Network
BERT	Bidirectional Encoder Representations from Transformers
MLP	Multi-Layer Perceptron
CLIP	Contrastive Language–Image Pretraining
Q-FedAvg	Quality-Aware Federated Averaging
FedDQ	Federated Data Quality

References

1. Al-Haija, Q.A.; Droos, A. A comprehensive survey on deep learning-based intrusion detection systems in Internet of Things (IoT). *Expert Syst.* **2025**, *42*, e13726. [CrossRef]
2. Marine-Roig, E.; Clave, S.A. Tourism analytics with massive user-generated content: A case study of Barcelona. *J. Destin. Mark. Manag.* **2015**, *4*, 162–172. [CrossRef]
3. Govindankutty, S.; Gopalan, S.P. From fake reviews to fake news: A novel pandemic model of misinformation in digital networks. *J. Theor. Appl. Electron. Commer. Res.* **2023**, *18*, 1069–1085. [CrossRef]
4. Glenski, M.; Ayton, E.; Mendoza, J.; Volkova, S. Multilingual multimodal digital deception detection and disinformation spread across social platforms. *arXiv* **2019**, arXiv:1909.05838. [CrossRef]
5. Karpii, O.; Mykhailyk, N. *Strategies for Integrating Marketing and Digital Reputation Management in the Modern Online Business Environment*; Publishing House "Baltija Publishing": Riga, Latvia, 2025.
6. Siow, E.; Tiropanis, T.; Hall, W. Analytics for the internet of things: A survey. *ACM Comput. Surv. (CSUR)* **2018**, *51*, 1–36. [CrossRef]
7. Marjani, M.; Nasaruddin, F.; Gani, A.; Karim, A.; Hashem, I.A.T.; Siddiqa, A.; Yaqoob, I. Big IoT data analytics: Architecture, opportunities, and open research challenges. *IEEE Access* **2017**, *5*, 5247–5261. [CrossRef]
8. Indrani, L.; Gadiraju, D.; Baligodugula, V.V. Federated Learning: Recent Advances and Future Directions. *TechRxiv* **2025**. [CrossRef] [PubMed]
9. Dembani, R.; Karvelas, I.; Akbar, N.A.; Rizou, S.; Tegolo, D.; Fountas, S. Agricultural data privacy and federated learning: A review of challenges and opportunities. *Comput. Electron. Agric.* **2025**, *232*, 110048. [CrossRef]
10. Uddin, M.P.; Xiang, Y.; Hasan, M.; Bai, J.; Zhao, Y.; Gao, L. A Systematic Literature Review of Robust Federated Learning: Issues, Solutions, and Future Research Directions. *ACM Comput. Surv.* **2025**, *57*, 1–62. [CrossRef]
11. Zhang, H.; Jia, X.; Chen, C. Deep Learning-Based Real-Time Data Quality Assessment and Anomaly Detection for Large-Scale Distributed Data Streams. *Int. J. Med. All Body Health Res.* **2025**, *6*, 1–11 .
12. Zhao, L.; Hu, S.; Wang, Q.; Jiang, J.; Shen, C.; Luo, X.; Hu, P. Shielding collaborative learning: Mitigating poisoning attacks through client-side detection. *IEEE Trans. Dependable Secur. Comput.* **2020**, *18*, 2029–2041.
13. Xu, J.; Zhang, C.; Jin, L.; Su, C. Data Quality-Aware Federated Learning for Fake Review Detection. In Proceedings of the 2025 7th International Conference on Software Engineering and Computer Science (CSECS), Taicang, China, 21–23 March 2025; pp. 1–6.

14. Kairouz, P.; McMahan, H.B.; Avent, B.; Bellet, A.; Bennis, M.; Bhagoji, A.N.; Bonawitz, K.; Charles, Z.; Cormode, G.; Cummings, R.; et al. Advances and open problems in federated learning. *Found. Trends Mach. Learn.* **2021**, *14*, 1–210. [CrossRef]
15. Tan, A.Z.; Yu, H.; Cui, L.; Yang, Q. Towards personalized federated learning. *IEEE Trans. Neural Netw. Learn. Syst.* **2022**, *34*, 9587–9603. [CrossRef]
16. Pillutla, K.; Kakade, S.M.; Harchaoui, Z. Robust aggregation for federated learning. *IEEE Trans. Signal Process.* **2022**, *70*, 1142–1154. [CrossRef]
17. Zhu, B.; Wang, L.; Pang, Q.; Wang, S.; Jiao, J.; Song, D.; Jordan, M.I. Byzantine-robust federated learning with optimal statistical rates. In Proceedings of the International Conference on Artificial Intelligence and Statistics, PMLR, Valencia, Spain, 25–27 April 2023; pp. 3151–3178.
18. Duan, H.; Hu, Q.; Wang, J.; Yang, L.; Xu, Z.; Liu, L.; Min, X.; Cai, C.; Ye, T.; Zhang, X.; et al. Finevq: Fine-grained user generated content video quality assessment. In Proceedings of the Computer Vision and Pattern Recognition Conference, Nashville, TE, USA, 11–15 June 2025; pp. 3206–3217.
19. Martin, L.; Sanchez, L.; Lanza, J.; Sotres, P. Development and evaluation of Artificial Intelligence techniques for IoT data quality assessment and curation. *Internet Things* **2023**, *22*, 100779. [CrossRef]
20. Bahaa, M.; Hany, M.; Zakaria, E.E. Advancing Automated Deception Detection: A Multimodal Approach to Feature Extraction and Analysis. In *Proceedings of the International Conference on Intelligent Systems, Blockchain, and Communication Technologies*; Springer: Cham, Switzerland, 2024; pp. 727–738.
21. Allam, H.; Makubvure, L.; Gyamfi, B.; Graham, K.N.; Akinwolere, K. Text classification: How machine learning is revolutionizing text categorization. *Information* **2025**, *16*, 130. [CrossRef]
22. Chang, Y. Research on the authenticity evaluation and recognition of social media health communication information based on deep learning algorithms. In Proceedings of the International Conference on Image Processing, Machine Learning and Pattern Recognition, Guangzhou, China, 13–15 September 2024; pp. 364–369.
23. Romero, R.; Celard, P.; Sorribes-Fdez, J.M.; Vieira, A.S.; Iglesias, E.L.; Borrajo, L. MobyDeep: A lightweight CNN architecture to configure models for text classification. *Knowl.-Based Syst.* **2022**, *257*, 109914. [CrossRef]
24. Jagannatha, A.N.; Yu, H. Structured prediction models for RNN based sequence labeling in clinical text. In Proceedings of the Conference on Empirical Methods in Natural Language Processing, Conference on Empirical Methods in Natural Language Processing, Austin, TX, USA, 1–5 November 2016; Volume 2016, p. 856.
25. Gardazi, N.M.; Daud, A.; Malik, M.K.; Bukhari, A.; Alsahfi, T.; Alshemaimri, B. BERT applications in natural language processing: A review. *Artif. Intell. Rev.* **2025**, *58*, 166. [CrossRef]
26. Samadi, M.; Mousavian, M.; Momtazi, S. Deep contextualized text representation and learning for fake news detection. *Inf. Process. Manag.* **2021**, *58*, 102723. [CrossRef]
27. Zhou, L.; Burgoon, J.K.; Nunamaker, J.F.; Twitchell, D. Automating linguistics-based cues for detecting deception in text-based asynchronous computer-mediated communications. *Group Decis. Negot.* **2004**, *13*, 81–106.
28. Yao, Y.; Viswanath, B.; Cryan, J.; Zheng, H.; Zhao, B.Y. Automated crowdturfing attacks and defenses in online review systems. In Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, Dallas, TX, USA, 30 October–3 November 2017; pp. 1143–1158.
29. Cao, N.; Ji, S.; Chiu, D.K.; Gong, M. A deceptive reviews detection model: Separated training of multi-feature learning and classification. *Expert Syst. Appl.* **2022**, *187*, 115977. [CrossRef]
30. Zheng, T.; Lin, Z.; Zhang, Y.; Jiao, Q.; Su, T.; Tan, H.; Fan, Z.; Xu, D.; Law, R. Revisiting review helpfulness prediction: An advanced deep learning model with multimodal input from Yelp. *Int. J. Hosp. Manag.* **2023**, *114*, 103579. [CrossRef]
31. Abouelenien, M.; Pérez-Rosas, V.; Mihalcea, R.; Burzo, M. Deception detection using a multimodal approach. In Proceedings of the 16th International Conference on Multimodal Interaction, Istanbul, Turkey, 12–16 November 2014; pp. 58–65.
32. Brocardo, M.L.; Traore, I.; Saad, S.; Woungang, I. Verifying online user identity using stylometric analysis for short messages. *J. Netw.* **2014**, *9*, 3347. [CrossRef]
33. Mishchenko, L.; Klymenko, I. Method for detecting fake news through writing style. *Tech. Sci. Technol.* **2023**, *4*, 82–90. [CrossRef]
34. Kim, B.; Calin, D.; Tenny, N.; Shariat, M.; Fan, M. Device centric distributed compute, orchestration and networking. *IEEE Wirel. Commun.* **2023**, *30*, 6–8. [CrossRef]
35. Antunes, R.S.; Andre da Costa, C.; Kuderle, A.; Yari, I.A.; Eskofier, B. Federated learning for healthcare: Systematic review and architecture proposal. *ACM Trans. Intell. Syst. Technol. (TIST)* **2022**, *13*, 1–23. [CrossRef]
36. Verma, R.K.; Kishor, K.; Galletta, A. Federated Learning Shaping the Future of Smart City Infrastructure. In *Federated Learning for Smart Communication Using IoT Application*; Chapman and Hall/CRC: Boca Raton, FL, USA, 2024; pp. 196–216.
37. Silva, T.W.; Morais, D.C.; Andrade, H.G.; Lima, A.M.; Melcher, E.U.; Brito, A.V. Environment for integration of distributed heterogeneous computing systems. *J. Internet Serv. Appl.* **2018**, *9*, 4. [CrossRef]
38. Zhang, H.; Li, C.; Dai, W.; Zou, J.; Xiong, H. Federated Learning Based on Model Discrepancy and Variance Reduction. *IEEE Trans. Neural Netw. Learn. Syst.* **2025**, *36*, 10407–10421. [CrossRef] [PubMed]

39. Elkordy, A.R.; Avestimehr, A.S. HeteroSAg: Secure aggregation with heterogeneous quantization in federated learning. *IEEE Trans. Commun.* **2022**, *70*, 2372–2386. [CrossRef]
40. Wang, Z.; Zhang, Z.; Tian, Y.; Yang, Q.; Shan, H.; Wang, W.; Quek, T.Q. Asynchronous federated learning over wireless communication networks. *IEEE Trans. Wirel. Commun.* **2022**, *21*, 6961–6978. [CrossRef]
41. Tang, M.; Ning, X.; Wang, Y.; Sun, J.; Wang, Y.; Li, H.; Chen, Y. FedCor: Correlation-based active client selection strategy for heterogeneous federated learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 19–20 June 2022; pp. 10102–10111.
42. Wu, X.; Huang, F.; Hu, Z.; Huang, H. Faster adaptive federated learning. In Proceedings of the AAAI Conference on Artificial Intelligence, Washington, DC, USA, 7–14 February 2023; Volume 37, pp. 10379–10387.
43. Albasyoni, A.; Safaryan, M.; Condat, L.; Richtárik, P. Optimal gradient compression for distributed and federated learning. *arXiv* **2020**, arXiv:2010.03246. [CrossRef]
44. Latif, N.; Ma, W.; Ahmad, H.B. Advancements in securing federated learning with IDS: A comprehensive review of neural networks and feature engineering techniques for malicious client detection. *Artif. Intell. Rev.* **2025**, *58*, 91. [CrossRef]
45. Kabbaj, H.; El-Azouzi, R.; Kobbane, A. Robust federated learning via weighted median aggregation. In Proceedings of the 2024 2nd International Conference on Federated Learning Technologies and Applications (FLTA), Valencia, Spain, 17–20 September 2024; pp. 298–303.
46. Wang, T.; Zheng, Z.; Lin, F. Federated learning framework based on trimmed mean aggregation rules. *Expert Syst. Appl.* **2025**, *270*, 126354. [CrossRef]
47. Shyn, S.K.; Kim, D.; Kim, K. Fedccea: A practical approach of client contribution evaluation for federated learning. *arXiv* **2021**, arXiv:2106.02310.
48. Song, Z.; Sun, H.; Yang, H.H.; Wang, X.; Zhang, Y.; Quek, T.Q. Reputation-based federated learning for secure wireless networks. *IEEE Internet Things J.* **2021**, *9*, 1212–1226. [CrossRef]
49. Shao, Y.; Li, J.; Shi, L.; Wei, K.; Ding, M.; Li, Q.; Li, Z.; Chen, W.; Jin, S. Robust Model Aggregation for Heterogeneous Federated Learning: Analysis and Optimizations. *arXiv* **2024**, arXiv:2405.06993. [CrossRef]
50. Yin, D.; Chen, Y.; Kannan, R.; Bartlett, P. Byzantine-robust distributed learning: Towards optimal statistical rates. In Proceedings of the International Conference on Machine Learning, PMLR, Stockholm, Sweden, 10–15 July 2018; pp. 5650–5659.
51. Segal, A.; Marcedone, A.; Kreuter, B.; Ramage, D.; McMahan, H.B.; Seth, K.; Bonawitz, K.; Patel, S.; Ivanov, V. Practical secure aggregation for privacy-preserving machine learning. In Proceedings of the CCS '17: Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, Dallas, TX, USA, 30 October–3 November 2017.
52. Yildirim Okay, F.; Ozdemir, S.; Xiao, Y. Fog computing-based privacy preserving data aggregation protocols. *Trans. Emerg. Telecommun. Technol.* **2020**, *31*, e3900. [CrossRef]
53. Nguyen, D.C.; Ding, M.; Pham, Q.V.; Pathirana, P.N.; Le, L.B.; Seneviratne, A.; Li, J.; Niyato, D.; Poor, H.V. Federated learning meets blockchain in edge computing: Opportunities and challenges. *IEEE Internet Things J.* **2021**, *8*, 12806–12825. [CrossRef]
54. Kulynych, B.; Gomez, J.F.; Kaissis, G.; du Pin Calmon, F.; Troncoso, C. Attack-aware noise calibration for differential privacy. *Adv. Neural Inf. Process. Syst.* **2024**, *37*, 134868–134901.
55. Chen, Y.; Luo, F.; Li, T.; Xiang, T.; Liu, Z.; Li, J. A training-integrity privacy-preserving federated learning scheme with trusted execution environment. *Inf. Sci.* **2020**, *522*, 69–79. [CrossRef]
56. Zhang, C.; Xie, Y.; Bai, H.; Yu, B.; Li, W.; Gao, Y. A survey on federated learning. *Knowl.-Based Syst.* **2021**, *216*, 106775. [CrossRef]
57. Jayaraman, B.; Evans, D. Evaluating differentially private machine learning in practice. In Proceedings of the 28th USENIX Security Symposium (USENIX Security 19), Santa Clara, CA, USA, 14–16 August 2019; pp. 1895–1912.
58. Davis, R. Knowledge-based systems. *Science* **1986**, *231*, 957–963. [CrossRef] [PubMed]
59. McMahan, B.; Moore, E.; Ramage, D.; Hampson, S.; y Arcas, B.A. Communication-efficient learning of deep networks from decentralized data. In Proceedings of the Artificial Intelligence and Statistics, PMLR, Lauderdale, FL, USA, 20–22 April 2017; pp. 1273–1282.
60. Li, T.; Sahu, A.K.; Zaheer, M.; Sanjabi, M.; Talwalkar, A.; Smith, V. Federated optimization in heterogeneous networks. *Proc. Mach. Learn. Syst.* **2020**, *2*, 429–450.
61. Song, S.; Li, Y.; Wan, J.; Fu, X.; Jiang, J. Data quality-aware client selection in heterogeneous federated learning. *Mathematics* **2024**, *12*, 3229. [CrossRef]
62. Wei, K.; Li, J.; Ding, M.; Ma, C.; Yang, H.H.; Farokhi, F.; Jin, S.; Quek, T.Q.; Poor, H.V. Federated learning with differential privacy: Algorithms and performance analysis. *IEEE Trans. Inf. Forensics Secur.* **2020**, *15*, 3454–3469. [CrossRef]
63. Lin, Y.M.; Gao, Y.; Gong, M.G.; Zhang, S.J.; Zhang, Y.Q.; Li, Z.Y. Federated learning on multimodal data: A comprehensive survey. *Mach. Intell. Res.* **2023**, *20*, 539–553. [CrossRef]

64. Sanh, V.; Debut, L.; Chaumond, J.; Wolf, T. DistilBERT, a distilled version of BERT: Smaller, faster, cheaper and lighter. *arXiv* **2019**, arXiv:1910.01108.
65. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.C. Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 4510–4520.