



## Full Length Article

## Blockchain-empowered cluster distillation federated learning for heterogeneous smart grids

Zhihao Zhou <sup>a,b</sup>, Yunhua He <sup>a,\*</sup>, Wei Zhang <sup>c</sup>, Zhaohui Ding <sup>c</sup>, Bin Wu <sup>d</sup>, Ke Xiao <sup>a</sup>

<sup>a</sup> School of Information Science and Technology, North China University of Technology, 100144, Beijing, China

<sup>b</sup> School of Cyberspace Science and Technology, Beijing Institute of Technology, 100864, Beijing, China

<sup>c</sup> China Datang Corporation Science and Technology Research Institute, 100043, Beijing, China

<sup>d</sup> Institute of Information Engineering, Chinese Academy of Science, 100864, Beijing, China

## ARTICLE INFO

**Keywords:**

Smart grid  
Blockchain  
Federated learning  
Device heterogeneity  
Knowledge distillation  
Quality assessment

## ABSTRACT

Smart grids have evolved into complex cyber-physical systems generating massive operational data that enable AI-driven optimization. While federated learning offers privacy preservation for sensitive energy data, its implementation in smart grid environments faces significant challenges from device heterogeneity and scalability limitations. Existing approaches inadequately address these challenges, particularly in handling computational disparities, non-independent and identically distributed data distributions, and quality control mechanisms simultaneously. This paper proposes a Blockchain-Empowered Cluster Distillation Federated Learning (BECDFL) framework that establishes a secure infrastructure through blockchain while addressing heterogeneity by categorizing participants according to computing capabilities and data distribution clustering. For resource-constrained devices, knowledge distillation enables effective model training without excessive computational demands. Additionally, quality assessment evaluates both data and model contributions, ensuring reliable performance in large-scale deployments. Experimental evaluations of the prototype system demonstrate that BECDFL achieves superior performance in effectiveness, robustness, and system sustainability compared to conventional federated learning approaches. The proposed solution offers a comprehensive framework for privacy-preserving machine learning in heterogeneous smart grid environments while maintaining high-quality model performance despite resource constraints and data variations.

### 1. Introduction

Smart grids constitute a paradigmatic transformation in electric power infrastructure, facilitating real-time monitoring, bidirectional communication protocols, and autonomous energy optimization through the integration of advanced digital technologies with conventional power grids [1]. The European Union had deployed approximately 223 million smart meters by the end of 2024, covering 77% of consumers through a substantial capital commitment of €47 billion [2]. Concurrently, power distribution grids across the EU27 + UK require estimated modernization and digitalization investment of €145-170 billion for 2020-2030 [3]. This infrastructure expansion generates unprecedented data volumes, with each million smart meters producing approximately 3TB of energy consumption data annually [4]. Consequently, artificial intelligence technologies have become essential for critical functions such as demand forecasting, fault detection, and system optimization [5]. Given the sensitive nature of energy consump-

tion data, federated learning (FL) has emerged as a promising paradigm that enables collaborative model training without centralizing private data [6-8]. This approach addresses regulatory compliance requirements while maintaining the benefits of distributed intelligence across the grid ecosystem [9,10].

However, federated learning implementation encounters two critical challenges within smart grid environments: heterogeneity and scalability constraints. First, heterogeneity manifests at both the hardware level in terms of computational capabilities and storage capacity and the data level in terms of distribution characteristics and scale. Hardware disparities range from resource-constrained smart meters to powerful substation servers. Concurrently, pervasive non-independent and identically distributed (Non-IID) data patterns fundamentally compromise training efficacy [11]. Second, scalability challenges emerge, as contemporary distribution networks must orchestrate participating devices while managing communication overheads that escalate with the number of participants [12]. The expansion of smart grid infrastructure

\* Corresponding author.

E-mail addresses: [cnedchou@gmail.com](mailto:cnedchou@gmail.com) (Z. Zhou), [heyunhua@ncut.edu.cn](mailto:heyunhua@ncut.edu.cn) (Y. He), [zhangwei@cdt-kxjs.com](mailto:zhangwei@cdt-kxjs.com) (W. Zhang), [dingchaohui@cdt-kxjs.com](mailto:dingchaohui@cdt-kxjs.com) (Z. Ding), [wubin@iie.ac.cn](mailto:wubin@iie.ac.cn) (B. Wu), [xiaoke@ncut.edu.cn](mailto:xiaoke@ncut.edu.cn) (K. Xiao).

and the integration of large-scale intelligent terminal devices have further increased the complexity of power grid management [13]. Quality assurance further compounds this predicament, as participating nodes may engage in fraudulent model submissions or malicious interference, thereby undermining federated learning efficacy.

Existing federated learning research has made progress in addressing these challenges, though critical limitations persist. Current approaches primarily address data distribution heterogeneity while neglecting device heterogeneity, a significant constraint in smart grid environments ranging from resource-constrained smart meters to computationally robust substations. While frameworks such as HeteroFL [14], personalized federated learning [15], and clustered approaches [16,17] have demonstrated efficacy in heterogeneous environments, computational resource heterogeneity remains inadequately addressed [18]. Moreover, these existing solutions typically exhibit insufficient security mechanisms and suboptimal generalization for complex power system dynamics. Furthermore, although recent advances in knowledge distillation [19,20] and adaptive aggregation [21] have enhanced federated learning efficiency, the generalization performance of the models still requires improvement. These methodologies lack systematic quality assessment frameworks, hindering effective identification and management of low-quality model contributions, thereby compromising critical smart grid functionalities including load forecasting, fault detection, and demand response management.

To tackle the above challenges, this paper proposes a Blockchain-Empowered Cluster Distillation Federated Learning (BECDFL) framework. This integrated approach leverages blockchain to establish a secure, transparent infrastructure for federated learning operations while addressing the multifaceted challenges present in smart grid environments. The blockchain component ensures tamper-proof model updates and secure transactions, maintaining the integrity of sensitive operational grid data. The framework systematically categorizes participants by computing power, and clusters them according to data characteristics. For devices with limited computational resources, particularly edge devices deployed throughout the distribution network, a knowledge distillation mechanism is incorporated, enabling effective model training without requiring extensive computational capabilities. Furthermore, a quality assessment mechanism is introduced to ensure reliable model quality in large-scale deployments by evaluating both data and model contributions, enhancing the accuracy of critical smart grid applications. The main contributions of this research are as follows:

- 1) We propose the Blockchain-Empowered Cluster Distillation Federated Learning framework. The framework uses blockchain to ensure security and trust, while clustering devices based on computing power and data characteristics to handle device heterogeneity.
- 2) We design a distillation-driven quality-aware hierarchical aggregation mechanism where devices with limited computing resources learn from more powerful devices through knowledge distillation. The mechanism assesses model and data quality using cosine similarity and marginal loss metrics to determine aggregation weights, improving the training quality for resource-constrained devices in complex model scenarios.
- 3) We construct a prototype system and conduct comprehensive experimental evaluations. The results demonstrate that the proposed scheme exhibits advantages in terms of effectiveness, robustness, and system sustainability.

This paper follows this organizational structure: [Section 2](#) outlines the related work. [Section 3](#) describes the system model as well as the design goals. [Section 4](#) details the blockchain-empowered cluster distillation federated learning with quality assessment methodology. [Section 5](#) presents the system implementation and performance evaluation results. [Section 6](#) concludes the paper.

## 2. Related works

[Table 1](#) summarizes research in federated learning approaches for smart grid environments. Although existing methods have made progress, they remain insufficient in simultaneously addressing computational resource variations and ensuring model quality. These limitations highlight the importance of developing integrated frameworks that can comprehensively tackle heterogeneity challenges while maintaining model reliability.

### 2.1. Heterogeneity-aware FL in smart grids

To address the heterogeneity challenges in the smart grid environments, Wen et al. [14] designed the HeteroFL framework, which employs prototype-based class imbalance learning methods to address heterogeneous data distributions in power theft detection. Fan et al. [15] developed three personalized strategy models for building energy management systems characterized by data scarcity and heterogeneity. However, these approaches generally exhibit inadequate security mechanisms, making them vulnerable to malicious attacks and data tampering threats. Cluster-based federated learning, which groups clients according to data distribution characteristics, has demonstrated advantages in heterogeneous environments. Lu et al. [16] focus on clustering extracted data features, aggregating client nodes with similar data distributions for federated learning. The energy and distribution-aware collaborative clustering algorithm proposed by Lee and Ko [17] advanced this approach by balancing energy efficiency and learning performance in the Internet of Things environments through a hierarchical parameter aggregation mechanism based on device energy levels, physical distance, and data distribution characteristics. However, existing schemes insufficiently address computational resource heterogeneity, and the generalization performance of models requires further improvement.

Wang et al. [18] addressed this limitation by combining federated learning with transfer learning for heating, ventilation, and air conditioning (HVAC) systems, grouping buildings by type and facilitating knowledge transfer between different building categories to overcome data insufficiency. Zhao et al. [22] proposed the Federated learning algorithm based on Client Sampling and Gradient Projection (FedCSGP) that uses client sampling and gradient projection to resolve internal and external gradient conflicts, improving performance for resource-constrained clients. Wu and Xu [23] developed the Spatial-Temporal Adaptive Personalized Federated Learning (FedSTA) framework with adaptive local aggregation mechanisms to handle data heterogeneity across diverse energy systems. Despite these advancements, most existing solutions focus primarily on data distribution characteristics while neglecting computational capability variations among devices, lacking comprehensive frameworks that simultaneously address both types of heterogeneity.

### 2.2. FL quality and optimization

To improve the training efficiency and model performance of federated learning, Tong et al. [19] proposed FedMR, a federated learning framework based on mutual knowledge distillation that achieves bidirectional distillation transfer between global and local knowledge by introducing generator models at the server side. Xiao and Wu [20] propose a personalized federated learning framework for battery management scenarios that integrates clustering and knowledge distillation modules to enable knowledge transfer between different types of batteries. However, these methods lack systematic quality assessment mechanisms, making it difficult to identify and process low-quality model contributions, which cannot guarantee model reliability and system stability, thereby affecting their long-term stable operation in large-scale practical application scenarios.

**Table 1**

Comparison of federated learning approaches in heterogeneous smart grids.

Solution Category	Reference	Description	Shortcoming
Data Heterogeneity	[14,15]	Prototype-based class imbalance learning; Personalized strategy models for building energy management.	Weak security mechanisms; Vulnerability to attacks; Limited protection against data tampering.
Client Clustering	[16,17]	Client grouping by data distribution; Feature-based clustering; Energy and distribution-aware hierarchical aggregation.	Insufficient handling of computational resource heterogeneity; Limited model generalization performance.
Integration Strategies	[18,22,23]	FL with transfer learning for HVAC systems; Client sampling with gradient projection; Spatial-temporal adaptive aggregation.	Focus limited to data distribution; No comprehensive aggregation frameworks.
Quality Enhancement	[19–21,24]	Bidirectional knowledge distillation; Clustering with distillation for batteries; Selective layer aggregation; Adaptive learning with auto-retraining.	No systematic quality assessment; Poor low-quality contribution detection; Static distillation strategies; Disregard of device computational constraints.

Manzoor et al. [21] proposed the Adaptive Single Layer Aggregation framework, which reduces communication overhead and memory requirements by selectively aggregating network layers and incorporating quantization techniques. Abdulla et al. [24] combine adaptive learning with federated learning to create an energy consumption prediction framework that monitors prediction errors in real-time and automatically triggers model retraining when necessary, effectively reducing prediction errors and training time. Most approaches do not incorporate device computational capability as a core consideration in knowledge distillation design, resulting in a lack of specialized distillation mechanisms for computationally constrained devices. Furthermore, these methods typically employ static or simplistic distillation strategies that lack robust quality control mechanisms for maintaining model integrity.

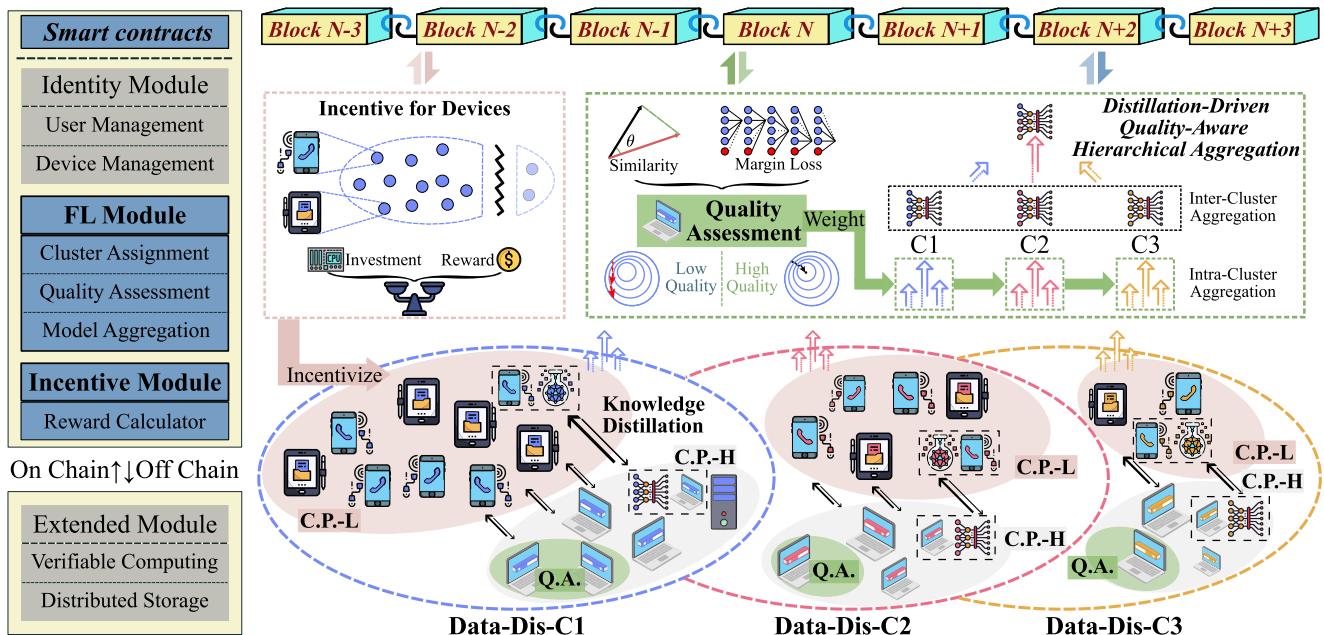
### 3. Blockchain-empowered cluster distillation federated learning framework

#### 3.1. Framework architecture and components

The proposed Blockchain-Empowered Cluster Distillation Federated Learning Framework, as it is illustrated in Fig. 1, establishes an integrated architecture optimized for effective federated learning in smart grid IoT environments. It aims to address critical challenges in smart

grid environments, including heterogeneous data distribution, computational resource constraints, and model quality assurance. The framework comprises the following constituent elements:

- **FL Blockchain:** The FL blockchain utilizes a consortium blockchain architecture to maintain an immutable and transparent record of system operations. The blockchain infrastructure documents device authentication, model updates, assessment results, and incentive metrics in a hierarchical node structure. Specifically, trusted institutional nodes such as governmental regulatory bodies and regional grid operators serve as full validator nodes participating in consensus operations, while high-capacity C.P.-H devices operate as light nodes. Additionally, the Extended Module provides capabilities for verifiable computing and distributed storage, enhancing the overall functionality and reliability of the blockchain system.
- **Smart Contract:** The modular smart contracts include essential modules for identity management, FL orchestration, and incentive mechanisms. The Identity Module manages device authentication and authorization processes. The FL Module orchestrates core federated learning operations, including quality assessment and model aggregation, with version control and validation procedures. The Incentive Module implements a game theory-based incentive, enabling the sys-

**Fig. 1.** Incentivized blockchain-empowered cluster distillation federated learning framework for smart grids.

tem to achieve collaborative security through a mixed-strategy Nash equilibrium.

- **FL Devices:** Diverse arrays of devices participate in the federated learning process. These devices are systematically categorized based on their computational capabilities into two primary classifications: computing-power high (C.P.-H) and computing-power low (C.P.-L). Despite their limited computational capacity, C.P.-L devices actively participate in the model training process and contribute to the model through collaborative knowledge distillation from their C.P.-H counterparts. Additionally, a subset of the C.P.-H devices is selected as quality assessment (Q.A) devices to complete the assigned assessment tasks.

Regarding the adaptation of blockchain to smart grid scenarios, the framework adopts the Practical Byzantine Fault Tolerance (PBFT) consensus mechanism due to its high transaction throughput and low latency, which is suitable for near real-time response requirements. In light of the inherent constraints on storage capacity and computational resources in blockchain systems, the framework implements an on-chain and off-chain paradigm for storage and computation.

The allocation between on-chain and off-chain collaborative storage is determined by factors such as data volume, access frequency, and verification requirements. On-chain storage is strictly limited to critical metadata including hashes of models, digital signatures, verification results, and transaction records, while off-chain storage handles high-volume data including complete model parameters, training datasets, and intermediate computational results. InterPlanetary File System (IPFS), as the off-chain decentralized storage solution, offers content-addressed storage with data integrity verification and efficient storage capabilities. Furthermore, given that smart contracts are ill-suited for executing complex computational tasks, business logic is decomposed according to computational complexity. Operations requiring O(1) or O(log n) complexity, such as authentication checks and signature verification are executed on-chain, while O(n) or higher complexity operations, such as model aggregation and quality assessment are performed off-chain. The off-chain computation employs verifiable computation techniques [25], and the results and verification proofs are subsequently recorded on the blockchain.

The incentive mechanism operates through the Incentive Module of smart contracts and establishes a game-theoretic framework to safeguard the system [26]. This mechanism is adapted from an approach employing model verifiers to achieve collaborative quality control in energy blockchain environments. Consequently, it is well-suited for the federated learning with quality control mechanisms within the smart grid discussed in this paper. The framework defines specific utility functions for participating devices and quality assessors, incorporating strategically designed penalty functions and dynamic incentive coefficients. Smart contracts execute algorithms that track malicious behaviors, calculate penalties, adjust incentive coefficients, and determine final rewards. This integrated approach promotes model quality and system integrity while maintaining computational efficiency suitable for resource-constrained IoT environments.

To enable efficient and secure interactions among system entities, the framework employs a **Distillation-Driven Quality-Aware Hierarchical Aggregation** mechanism as its core operational component. This component operates through several integrated processes. First, clients with similar data characteristics form clusters (represented as C1, C2, and C3), with clustering results recorded on the blockchain through a consensus-validated transaction, providing a basis for subsequent model aggregation and incentive distribution. Second, a knowledge distillation process enables effective collaborative knowledge transfer from complex models to simpler ones. Third, a quality assessment component ensures process integrity and reliability through an extended evaluation system by the Q.A. assessment. Finally, the hierarchical aggregation operates at two levels, as intra-cluster aggregation considers both data quality and

model performance, while inter-cluster aggregation accounts for cluster representativeness to generate the global model. A more detailed description of this process is provided in [Section 4](#).

Through this framework, blockchain provides a trustworthy infrastructure for federated learning while also creating an effective, secure, and quality-assured decentralized collaborative learning environment.

### 3.2. Design goals

Through the comprehensive framework and quality-aware cluster aggregation mechanism based on knowledge distillation, our approach effectively tackles the challenges of heterogeneity and scalability in smart grid federated learning environments. Specifically, the proposed approach aims to accomplish three primary design goals:

- **Effectiveness:** Our approach enables efficient global federated learning under heterogeneous data distributions and computational resource constraints.
- **Robustness:** Our approach is designed to maintain stable performance across diverse datasets, various data distributions, and in the presence of malicious clients, ensuring reliable model training.
- **Sustainability:** Our approach is designed to efficiently support large-scale device participation and frequent model updates while maintaining system performance and security guarantees.

## 4. Distillation-driven quality-aware hierarchical aggregation

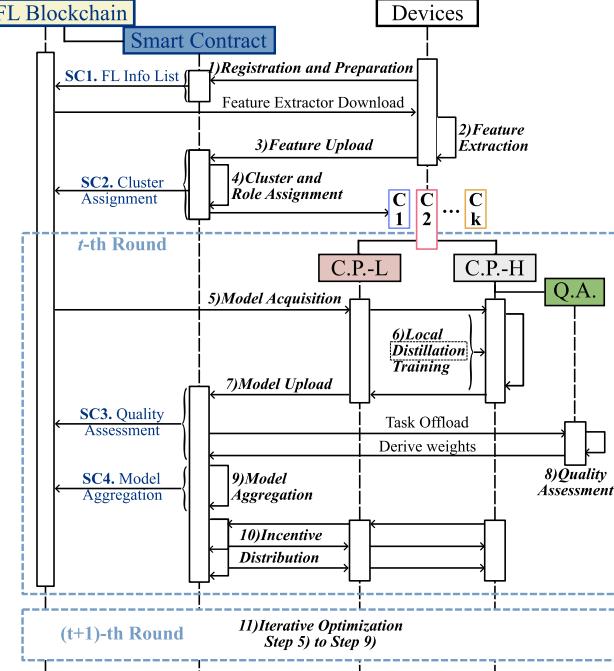
In the context of IoT-enabled smart grids characterized by diverse device types and imbalanced data distributions, this paper proposes a BECDFL framework. This framework leverages blockchain and smart contracts to create a unified approach to data security and federated learning that operates effectively while optimizing computational resources and mitigating the challenges posed by heterogeneous data distribution.

### 4.1. Detailed blockchain-empowered cluster distillation federated learning process

To enhance clarity and conciseness in process descriptions and representations, this paper uniformly refers to blockchain and its extended off-chain storage as FL Blockchain, while smart contracts and their extended off-chain computing are collectively termed Smart Contracts.

Specifically, the proposed blockchain-empowered cluster distillation federated learning method encompasses the following steps, as it is illustrated in [Fig. 2](#).

- 1) **Registration and Preparation.** Devices intending to participate in FL tasks must register their device information within the FL blockchain. Upon successful registration, devices become verified to apply for participation in FL tasks and acquire the encoder model.
- 2) **Feature Extraction.** Each device  $i$  conducts feature extraction on training dataset  $D_i$  to obtain the feature  $x_i$ .
- 3) **Feature Upload.** Each device  $i$  uploads the extracted feature vector  $x_i$  to the FL blockchain.
- 4) **Cluster and Role Assignment.** Upon receiving feature data from all participants, the smart contract assigns device  $i$  to the corresponding cluster  $C_k$ .
- 5) **Model Acquisition:** At training round  $t = 1$ , device  $i$  retrieves and downloads the initial model parameters  $\theta_g^{(t-1)}$  from the FL blockchain.
- 6) **Local Distillation Training.** Each device  $i$  updates its local model parameters  $\theta_i^{(t)}$  by applying knowledge distillation on its local training dataset  $D_i$ .
- 7) **Model Upload.** Each device  $i$  uploads its updated model parameters  $\theta_i^{(t)}$  to the FL blockchain.



**Fig. 2.** Cluster distillation federated learning process empowered by blockchain and smart contracts.

- 8) Quality Assessment. The quality assessment smart contract assesses the data and model quality to determine the aggregation weight  $w_i^{(t)}$  for device  $i$  after receiving updates from all participants in cluster  $C_k$ .
- 9) Model Aggregation. The smart contract executes global model aggregation based on the derived weights  $\alpha_i^{(t)}$ , yielding updated global model parameters  $\theta_g^{(t)}$ .
- 10) Iterative Optimization. Steps 5 through 9 are iteratively executed until the global model's loss converges, completing the training task.
- 11) Incentive Distribution. The incentive module of the smart contract computes and distributes rewards to devices participating in the federated learning process.

#### 4.2. Feature-based cluster and role assignment

In the device access and preparation phase, the BECDFL employs a blockchain-based decentralized identity authentication mechanism (DID) to record device identity information on the blockchain. Devices are required to declare their type  $Type_i$  and corresponding capability  $h_i$  during authentication and verification. Devices are categorized into computational devices and storage devices, with capabilities corresponding to computational power and storage capacity, respectively. Within the BECDFL framework, the FL blockchain maintains a list  $\mathbf{L}$  of devices participating in federated learning. For verified devices, the smart contract first checks whether their  $DID$  exists in list  $\mathbf{L}$ . For new DIDs, information regarding device type  $Type_i$  and capability  $h_i$  is recorded in  $\mathbf{L}$ ; for existing DIDs, the smart contract checks whether their capacity  $h_i$  requires updating and updates device information in  $\mathbf{L}$  accordingly. This process manages participating device identities and establishes the foundation for federated learning task execution, as detailed in [Algorithm 1](#).

After registration, devices obtain a specific autoencoder model  $\mathcal{F}_{enc} : \mathcal{X} \rightarrow \mathbb{R}^d$  from the FL blockchain. This model is pre-trained on a representative dataset that captures the target data characteristics, and it is periodically updated by high-capability devices to adapt to evolving data patterns. The lightweight and computationally efficient autoencoder extracts features from each device's local dataset  $D_i = \{(x_j, y_j)\}_{j=1}^{|D_i|}$ , gener-

---

#### Algorithm 1 Device FL info list maintenance.

```

Require: Verified device identity  $DID_i$ , Device type  $Type_i$ , Capability  $h_i$ 
Ensure: Updated Device List  $\mathbf{L}$ 
1: for each device  $DID_i$  do
2:   if  $DID_i$  does not exists in  $\mathbf{L}$  then
3:     add_to_list( $\mathbf{L}$ ,  $DID_i$ ,  $Type_i$ ,  $h_i$ )
4:   else
5:     if  $Type_i$  is new or  $h_i$  is new then
6:       update_capability( $\mathbf{L}$ ,  $DID_i$ ,  $Type_i$ ,  $h_i$ )
7:     end if
8:   end if
9: end for
10: return  $\mathbf{L}$ 

```

---

ating feature vectors  $\mathbf{x}_i \in \mathbb{R}^d$ , as shown in [Eq. \(1\)](#). The resulting feature vector  $\mathbf{x}_i$  provides a low-dimensional, condensed representation of the core distributional characteristics of the local dataset  $D_i$ , capturing key information that distinguishes different data distributions. When the autoencoder  $\mathcal{F}_{enc}$  projects data into a well-structured feature space, the centroid of these projections effectively represents the overall data distribution characteristics. This effective extraction and quantification of data distribution features serves as a prerequisite for achieving precise device clustering in heterogeneous federated learning environments.

$$\mathbf{X}_i = \frac{1}{|D_i|} \sum_{x_j \in D_i} \mathcal{F}_{enc}(x_j). \quad (1)$$

---

#### Algorithm 2 Cluster and role assignment.

```

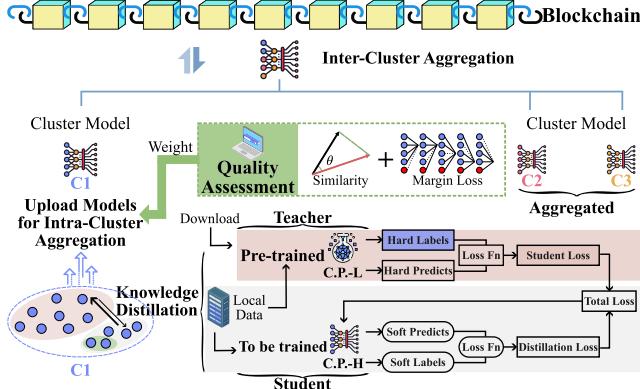
Require: Features  $\{\mathbf{x}_i\}_{i=1}^N$ , Number of clusters  $k$ , Capability  $h_i$ , Capability threshold  $ph$ 
Ensure: Cluster assignments and device roles
1:  $C \leftarrow KMeans(\{\mathbf{x}_i\}, k)$  // Perform clustering
2: for each cluster  $c \in C$  do
3:   for device  $i \in c$  do
4:     Assign role  $\leftarrow$  (if  $h_i \geq ph$  then C.P.-H else C.P.-L)
5:   end for
6: end for
7: return  $C$  and device roles

```

---

The device then uploads the extracted local data feature vectors  $\mathbf{x}_i$  to the FL blockchain. Upon receiving features from all participating devices, the off-chain computing module employs clustering algorithms to partition the devices into  $k$  clusters  $C = \{C_1, C_2, \dots, C_k\}$  based on data distribution. Compared to methods such as hierarchical clustering [27], DBSCAN [28], and Gaussian Mixture Models [29], the K-means algorithm is selected for clustering on large-scale datasets owing to its simple implementation and high computational efficiency. This makes it particularly suitable for federated learning scenarios that require rapid partitioning of homogeneous groups. The K-means algorithm operates through an iterative optimization process designed to minimize the sum of squared Euclidean distances between each device's feature vector and its assigned cluster centroid, thereby partitioning devices based on the similarity of their local data distributions.

Further, the system categorizes devices within each cluster into C.P.-H and C.P.-L groups to execute distinct federated learning tasks based on their processing capabilities. Devices with capability values exceeding a predefined threshold ( $ph$ ) are classified as C.P.-H and assigned computation-intensive tasks, including complex model training and parameter aggregation. A subset of selected C.P.-H devices are additionally assigned comprehensive quality assessment (Q.A.) responsibilities. In contrast, C.P.-L devices handle less demanding computational workloads. This capability-based differentiation enables C.P.-H devices to process larger-scale operations, execute complex algorithmic proce-



**Fig. 3.** Knowledge distillation federated learning with quality-aware cluster aggregation.

dures, and maintain federated learning stability under resource constraints. This hierarchical role assignment mechanism optimizes computational task allocation according to device capabilities, enhances overall resource utilization efficiency, and ensures the reliability of the federated learning process. The operational logic and implementation details of this cluster and role assignment approach are presented in [Algorithm 2](#).

#### 4.3. Role-specific distillation training for devices

The knowledge distillation local training and quality-aware cluster aggregation are shown in [Fig. 3](#), a process that reduces model complexity, enables efficient knowledge transfer between teacher and student models, and facilitates effective handling of heterogeneous data distributions.

The teacher model  $\mathcal{M}_\phi^T$  serves a critical role in the distillation process for high-quality knowledge transfer. The teacher model is deliberately designed to be more complex and expressive than the student models, incorporating a larger number of parameters that enable it to capture data patterns essential for accurate predictions. To ensure optimal performance, the teacher model is maintained and periodically updated through a specialized federated learning process that exclusively involves the C.P.-H devices.

During the  $t$ -th round of local training, computationally constrained client devices (C.P.-L) employ knowledge distillation to train lightweight student models  $\mathcal{M}_{\theta_i^{(t)}}^S$ , where  $\theta_i^{(t)}$  represents model parameters. An offline DIST[30] distillation approach with loss function  $\mathcal{L}_D$  is employed, wherein each student client device  $i$  generates soft labels  $y_j^s = \mathcal{M}_\phi^T(x_j)$  for its local dataset using the teacher model  $\mathcal{M}_\phi^T$ . Compared to traditional knowledge distillation methods, this approach focuses more on capturing the relative relationships of teacher model predictions rather than absolute probability values, which offers significant advantages in heterogeneous federated learning environments.

Under this framework, the distillation loss function  $\mathcal{L}_D$  is decomposed into inter-class  $\mathcal{L}_{\text{inter}}$  and intra-class  $\mathcal{L}_{\text{intra}}$  relational losses. Both of which can be expressed using the unified formulation  $\mathcal{L}_{\text{rel}}(\phi)$  given in [Eq. \(2\)](#),

$$\mathcal{L}_{\text{rel}}(\phi) := \frac{1}{|\Omega_\phi|} \sum_{k \in \Omega_\phi} d_p(\mathbf{Y}_{\phi(k)}^{(s)}, \mathbf{Y}_{\phi(k)}^{(t)}), \quad (2)$$

where  $\phi(k)$  is a mapping that specifies the slicing operation based on the relational type,  $d_p(\cdot, \cdot)$  is metric defined based on the Pearson correlation coefficient,  $\mathbf{Y}^{(s)}$  and  $\mathbf{Y}^{(t)}$  represent the prediction probability matrices of the teacher and student models, respectively, and  $\Omega_\phi$  denotes the index set. Specifically,  $\phi_{\text{inter}}(k) = (k, :)$  represents extracting the prediction distribution of the  $k$ -th sample across all classes, and  $\phi_{\text{intra}}(k) = (:, k)$  extracts the prediction probabilities of all samples

for the  $k$ -th class. When  $\phi = \phi_{\text{inter}}$ , the inter-class relational loss  $\mathcal{L}_{\text{inter}}$  quantifies prediction distribution discrepancies for each sample across the class space, thereby preserving discriminative boundaries between classes. Conversely, when  $\phi = \phi_{\text{intra}}$ , the intra-class relational loss  $\mathcal{L}_{\text{intra}}$  measures variations in prediction patterns among all samples for each class, maintaining consistency in intra-class predictions.

Model optimization combines these soft labels  $y_j^s$  with local hard labels  $y_j$ . The loss function expressed as shown in [Eq. \(3\)](#),

$$\mathcal{L}_i^{(t)} = \lambda_D \cdot \mathcal{L}_D(\mathcal{M}_{\theta_i^{(t)}}^S(x_j), y_j^s) + \lambda_C \cdot \mathcal{L}_C(\mathcal{M}_{\theta_i^{(t)}}^S(x_j), y_j), \quad (3)$$

where  $\mathcal{L}_D$  represents the distillation loss function, measuring similarity between student model outputs and soft labels. Weight coefficients  $\lambda_C, \lambda_D$  satisfy  $\lambda_C + \lambda_D = 1$ .  $\mathcal{L}_E$  denotes the cross-entropy loss function, calculating discrepancies between student model outputs and hard labels, as shown in [Eq. \(4\)](#),

$$\mathcal{L}_C(\mathcal{M}_{\theta_i^{(t)}}^S(x_j), y_j) = - \sum_{c=1}^C \mathbb{I}(y_j = c) \log \mathcal{M}_{\theta_i^{(t)}}^S(x_j)_c, \quad (4)$$

where  $\mathbb{I}(\cdot)$  is the indicator function,  $C$  represents the number of classes, and  $\mathcal{M}_{\theta_i^{(t)}}^S(x_j)_c$  denotes the student model's prediction probability for class  $c$ .

Based on the above content, the overall training objective can be formulated as [Eq. \(5\)](#). This knowledge distillation framework enables devices with limited computational resources to effectively participate in the federated learning process while maintaining high model performance and generalization capability, providing a robust theoretical framework for heterogeneous federated learning environments.

$$\mathcal{L}_i^{(t)} = \lambda_D(\beta \mathcal{L}_{\text{inter}} + \gamma \mathcal{L}_{\text{intra}}) + \lambda_C \mathcal{L}_C(\mathcal{M}_{\theta_i^{(t)}}^S(x_j), y_j). \quad (5)$$

#### 4.4. Hierarchical aggregation with quality assessment

The quality assessment module evaluates both data quality ( $d_i^t$ ) and model quality ( $m_i^t$ ) to generate weight parameters ( $w_i^t$ ) for the model aggregation process in BECDFL. This assessment addresses two critical challenges in heterogeneous IoT environments: varying data quality across devices and potential adversarial or erroneous model updates. The quality assessment procedure is detailed in [Algorithm 3](#). Specifi-

---

#### Algorithm 3 Quality assessment.

**Require:** Local models  $\{\theta_i^{(t)}\}_{i \in C_k}$ , Parameters  $\lambda, \gamma_m, \gamma_d$   
**Ensure:** Cluster aggregation weights  $\{w_i^{(t)}\}_{i \in C_k}$

- 1: Calculate uniformly aggregated global model  $\bar{\theta}_g^{(t)}$  according to [Eq. \(6\)](#)
- 2: **for** each device  $i \in C_k$  **do**
- 3:   Calculate self-excluded uniformly aggregated global model  $\bar{\theta}_{g-i}^{(t)}$  according to [Eq. \(7\)](#)
- 4:   Calculate cosine similarity  $\cos_i^{(t)}$  according to [Eq. \(8\)](#)
- 5:   Calculate marginal loss  $\delta_i^{(t)}$  according to [Eq. \(10\)](#)
- 6:   Calculate quality metric  $q_i^{(t)}$  according to [Eq. \(11\)](#)
- 7:   Calculate cluster aggregation weight  $w_i^{(t)}$  according to [Eq. \(12\)](#)
- 8: **end for**
- 9: **return**  $\{w_i^{(t)}\}_{i \in C_k}$

---

cally, on the  $t$ -th iteration, the quality assessment module first finds the uniformly aggregated global model  $\bar{\theta}_g^{(t)}$  using the local model parameters  $\theta_i^{(t)}$  of each device  $i$  in cluster  $C_k$ :

$$\bar{\theta}_g^{(t)} = \frac{1}{|C_k|} \sum_{i \in C_k} \theta_i^{(t)}. \quad (6)$$

Then, it computes the uniformly aggregated global model  $\bar{\theta}_{g-i}^{(t)}$  by aggregating all models except the model of device  $i$ :

$$\bar{\theta}_{g-i}^{(t)} = \frac{1}{|C_k| - 1} \sum_{j \in C_k \setminus \{i\}} \theta_j^{(t)}. \quad (7)$$

To eliminate the influence of device  $i$ 's own training model, the system evaluates the consistency between device  $i$ 's updates and the overall updates using cosine similarity. Cosine similarity is chosen for its magnitude-invariant measure of directional similarity, making it particularly suitable for federated learning scenarios where update magnitudes can vary significantly across devices without necessarily indicating quality differences, thereby accurately reflecting their actual contribution to the global model.

Based on this, the cosine similarity  $\cos_i^{(t)}$  between device  $i$ 's model parameters  $\theta_i^{(t)}$  and the global model  $\bar{\theta}_{g-i}^{(t)}$  (excluding device  $i$  itself) is calculated as:

$$\cos_i^{(t)} = \frac{\langle \theta_i^{(t)}, \bar{\theta}_{g-i}^{(t)} \rangle}{\|\theta_i^{(t)}\| \cdot \|\bar{\theta}_{g-i}^{(t)}\|}, \quad (8)$$

where  $\langle \cdot, \cdot \rangle$  denotes the inner product, and  $\|\cdot\|$  represents the Euclidean norm. Here, a normalization function  $Q(\mu_i, \lambda)$  is introduced, which incorporates two parameters: the evaluation metric  $\mu_i$  and a modulating hyperparameter  $\lambda$ . As shown in Eq. (9), this function regulates the differentiation among various metrics through exponential operations and employs normalization to ensure the sum of evaluation values equals 1 for subsequent weight allocation.

$$Q(\mu_i, \lambda, S_k) = \frac{\exp(\lambda\mu_i)}{\sum_{j \in S_k} \exp(\lambda\mu_j)}, \quad (9)$$

Based on the above function, the model quality  $m_i^t$  can be calculated as  $m_i^t = Q(\cos_i^{(t)}, \gamma_m, S_k)$ .

Furthermore, the data quality  $d_i^t$  of device  $i$  can be evaluated by calculating its marginal loss contribution  $\delta_i^{(t)}$  to the global model, as shown in Eq. (10). Marginal loss quantifies the change in the global model's overall loss when device  $i$ 's contribution is excluded from the current aggregation round, thereby capturing both the uniqueness and relevance of the device's data distribution. A larger marginal loss value  $\delta_i^{(t)}$  indicates that device  $i$ 's local data and corresponding model updates are more crucial in reducing the global model's loss function, thereby enhancing accuracy and generalization performance.

$$\delta_i^{(t)} = \ell_{-i}^{(t)} - \ell^{(t)}, \quad (10)$$

where  $\ell^{(t)}$  and  $\ell_{-i}^{(t)}$  represent the loss values of the global model  $\bar{\theta}_g^{(t)}$  and the global model  $\bar{\theta}_{g-i}^{(t)}$  excluding participant  $i$ , respectively. This suggests that either client  $i$  either possesses higher-quality data or has a data distribution that plays a more significant role in model improvement. Accordingly, the formula for data quality  $d_i^t$  is given by  $d_i^t = Q(\delta_i^{(t)}, \gamma_d, C_k)$ .

Finally, the comprehensive quality metric  $q_i^{(t)}$  is defined based on the data quality  $d_i^{(t)}$  and model quality  $m_i^{(t)}$ , which are derived from marginal loss and cosine similarity, as defined in Eq. (11). In constructing the comprehensive quality metric  $q_i^{(t)}$ , the weight coefficient  $\gamma \in [0, 1]$  balances the relative importance of data quality  $d_i^{(t)}$  and model quality  $m_i^{(t)}$  in assessing each device's overall contributions. The optimal value of  $\gamma$  should be determined based on specific application requirements, the degree of data heterogeneity, and desired performance objectives. Setting  $\gamma = 0.5$  represents a balanced approach that treats data quality and model quality as equally important components.

$$q_i^{(t)} = \gamma d_i^{(t)} + (1 - \gamma) m_i^{(t)}. \quad (11)$$

The cluster aggregation weight  $w_i^{(t)}$  for participant  $i$  in round  $t$  is thus calculated as in Eq. (12).

$$w_i^{(t)} = \frac{q_i^{(t)}}{\sum_{j \in C_k} q_j^{(t)}}. \quad (12)$$

After completing the  $t$ -th round of training, the model aggregation process adopts a hierarchical strategy, first performing weighted aggregation within clusters, followed by uniform aggregation across clusters as described in Algorithm 4.

$$\mathcal{M}_{C_k}^S = \sum_{i \in C_k} w_i^{(t)} \mathcal{M}_{\theta_i^{(t)}}^S, \quad \mathcal{M}_{g^{(t)}}^S = \sum_{C_k \in C} \frac{1}{|C|} \mathcal{M}_{C_k}^S. \quad (13)$$

This aggregation mechanism effectively addresses data distribution heterogeneity and computational capability disparities, enhancing system scalability and processing efficiency.

#### Algorithm 4 Cluster model aggregation.

---

**Require:** Local models  $\{\theta_i^{(t)}\}_{i=1}^N$ , weights  $\{w_i^{(t)}\}_{i=1}^N$ , clusters  $\{C_k\}_{k=1}^K$   
**Ensure:** Global model  $\mathcal{M}_{g^{(t)}}^S$

```

1: for each cluster  $C_k \in \{C_1, \dots, C_K\}$  do
2:    $\mathcal{M}_{C_k}^S \leftarrow 0$ 
3:   for each device  $i \in C_k$  do
4:      $\mathcal{M}_{C_k}^S \leftarrow \mathcal{M}_{C_k}^S + w_i^{(t)} \mathcal{M}_{\theta_i^{(t)}}^S$ 
5:   end for
6: end for
7:  $\mathcal{M}_{g^{(t)}}^S \leftarrow 0$ 
8: for each cluster  $C_k \in C$  do
9:    $\mathcal{M}_{g^{(t)}}^S \leftarrow \mathcal{M}_{g^{(t)}}^S + \frac{1}{|C|} \mathcal{M}_{C_k}^S$ 
10: end for
11: return  $\mathcal{M}_{g^{(t)}}^S$ 

```

---

#### 4.5. Innovation and convergence analysis

The proposed BECDFL framework demonstrates architectural and methodological distinctions from existing paradigms. While Cluster-based FL relies on client clustering with centralized trust assumptions, and Blockchain-based FL provides decentralized infrastructure without addressing heterogeneity, BECDFL combines blockchain-empowered cluster distillation in a integrated architecture, as it is shown in Table 2.

Our key innovation lies in the quality assessment mechanism that employs a dual-metric approach for comprehensive evaluation, with marginal loss for data quality assessment and cosine similarity for model quality assessment. This dual-quality assessment integrated with decentralized blockchain verification represents a fundamental advancement beyond existing federated learning paradigms.

As summarized in Table 3, the client-side computational burden is identical for all methods at  $\mathcal{O}(E|D_i|C_{\text{model}})$ . The difference is implemented at the server level, where the proposed BECDFL transforms the aggregation process from a simple weighted average into an quality-driven synthesis. BECDFL introduces a two-stage hierarchical aggregation  $\mathcal{O}(md + Kd)$  and quality assessment with cosine similarity check ( $\mathcal{O}(md)$ ) and a marginal loss evaluation ( $\mathcal{O}(mB_k C_{\text{fwd}})$ ), enabling measurement of the value of each update. BECDFL makes a trade-off, accepting a modest and linear increase in server-side complexity in exchange for enhanced system robustness. This investment in computation and memory equips the system with robustness against low-quality contributions and a mechanism for managing client diversity.

Further, We establish non-convex convergence guarantees for the aggregation of models trained by C.P.-L devices, with teacher models trained by C.P.-H devices providing soft labels.

**Definition 1** (L-Smooth Function). A differentiable function  $F : \mathbb{R}^d \rightarrow \mathbb{R}$  is  $L$ -smooth if  $F(\theta_1) \leq F(\theta_2) + \langle \nabla F(\theta_2), \theta_1 - \theta_2 \rangle + \frac{L}{2} \|\theta_1 - \theta_2\|^2$  for all  $\theta_1, \theta_2$  in its domain.

**Definition 2** (Local Objective). For device  $i$ , the objective incorporating distillation is  $F_i(\theta) := (1 - \lambda_{D,i})F_i(\theta) + \lambda_{D,i}D_i(\theta; \phi)$ , where  $F_i$  corresponds to the hard-label loss and  $D_i(\cdot; \phi)$  to the distillation loss.

**Table 2**  
Architecture and process of BECDFL in contrast to existing paradigms.

	Cluster-based FL	Blockchain-based FL	BECDFL
<b>Heterogeneity Solution</b>	Client Clustering	Not Addressed	Client Clustering + Knowledge Distillation
<b>Device Diversity</b>	Assumes Homogeneity	Not Addressed	Roles Specific
<b>Trust Model</b>	Centralized Trusted Server	Decentralized Ledger	On-Chain Verifiable Process
<b>Robustness Mechanism</b>	Basic	Basic	Quality Assessment
<b>Architecture</b>	Monolithic	Loosely Coupled	Tightly Integrated

**Table 3**  
Core time and space complexity comparison per round.

Aspect	BECDFL	FedAvg [31]/ FedProx[32]
<b>Client Computation</b>	Dominated by local training cost: $\mathcal{O}(E D_i C_{\text{model}})$ per client.	
<b>Server Computation</b>	<b>1. Two-stage Aggregation:</b> $\mathcal{O}(md + Kd)$	
<b>2. Quality Assessment:</b>		
- Cosine Sim: $\mathcal{O}(md)$		
- Marginal Loss: $\mathcal{O}(mB_k C_{\text{fwd}})$	<b>Simple Aggregation:</b> $\mathcal{O}(md)$	
<b>Server Space</b>	Base model: $\mathcal{O}(d)$	
+ <b>State storage:</b> $\mathcal{O}(Nd_f + Tmd)$	Base model: $\mathcal{O}(d)$	

**Notation.**  $m$ : participants per round;  $d$ : model dim.;  $E$ : local epochs/steps;  $|D_i|$ : client dataset size;  $C_{\text{model}}$ : per-sample training cost;  $K$ : #clusters;  $B_k$ : validation batch size;  $C_{\text{fwd}}$ : per-sample forward pass cost;  $N$ : total devices;  $T$ : total rounds;  $d_f$ : feature dim.

**Definition 3** (Global Objective). Let  $S_L$  be the C.P.-L index set,  $C_k^L := C_k \cap S_L$ , and  $\mathcal{K}_L := \{k : C_k^L \neq \emptyset\}$ . With the averaged intra-cluster quality weights  $\bar{w}_i$  and the inter-cluster weights  $\bar{v}_k$ , define  $F_{C_k^L}^l(\theta) := \sum_{i \in C_k^L} \bar{w}_i F_i^l(\theta)$  and  $\tilde{F}_L(\theta) := \sum_{k \in \mathcal{K}_L} \bar{v}_k F_{C_k^L}^l(\theta)$ .

**Definition 4** (Level Set). Given an initial point  $\theta^{(0)}$ , define the level set  $\mathcal{L}_0 := \{\theta : \tilde{F}_L(\theta) \leq \tilde{F}_L(\theta^{(0)})\}$ .

**Definition 5** (Weighted Heterogeneity Measures). Define

$$\epsilon_{\text{intra}}^2 := \sup_{\theta \in \mathcal{L}_0} \sum_k \bar{v}_k \sum_{i \in C_k^L} \bar{w}_i \left\| \nabla F_i^l(\theta) - \nabla F_{C_k^L}^l(\theta) \right\|^2, \quad (14)$$

$$\epsilon_{\text{inter}}^2 := \sup_{\theta \in \mathcal{L}_0} \sum_k \bar{v}_k \left\| \nabla F_{C_k^L}^l(\theta) - \nabla \tilde{F}_L(\theta) \right\|^2. \quad (15)$$

**Assumption 1** (Problem Smoothness and Lower Boundedness).  $\tilde{F}_L$  is lower bounded and  $L$ -smooth on  $\mathcal{L}_0$  in the sense of **Definition 1**.

**Assumption 2** (Unbiased SGD). For each device, stochastic gradients are unbiased for  $F_i^l$  and satisfy  $\mathbb{E}\|g_i(\theta) - \nabla F_i^l(\theta)\|^2 \leq \sigma^2$  and  $\mathbb{E}\|\nabla F_i^l(\theta)\|^2 \leq G^2$ .

**Assumption 3** (Bounded Heterogeneity). The measures in **Definition 5** are finite:  $\epsilon_{\text{intra}}^2 < \infty$  and  $\epsilon_{\text{inter}}^2 < \infty$ .

**Assumption 4** (Bounded Distillation Error). Let  $\tilde{\mathcal{V}}\mathcal{L}_i(\theta)$  be the actual per-step update that mixes hard and soft labels. Then  $\mathbb{E}\|\tilde{\mathcal{V}}\mathcal{L}_i(\theta) - \nabla F_i^l(\theta)\|^2 \leq B^2$ .

**Assumption 5** (Bounded Aggregation Weights). Weights satisfy  $0 < \bar{w}_i \leq w_{\max} \leq 1$  with  $\sum_{i \in C_k^L} \bar{w}_i = 1$  and  $\bar{v}_k \geq 0$  with  $\sum_k \bar{v}_k = 1$ . Online weights are bounded and conditionally aligned:  $\mathbb{E}[w_i^{(t)} | \theta^{(t)}] = \bar{w}_i$ ,  $\mathbb{E}[v_k^{(t)} | \theta^{(t)}] = \bar{v}_k$ , with bounded conditional variances.

**Assumption 6** (Local Steps and StepSize). Each device performs  $E \geq 1$  local SGD steps with stepsize  $\eta \leq c/(LE)$  for an absolute constant  $c \in (0, 1)$ .

**Theorem 1** (Non-Convex Convergence). Under **Assumptions 1–6** and with C.P.-L devices participating, let  $\{\theta^{(t)}\}_{t=0}^T$  be the global iterates produced by the aggregation

$$\theta_{C_k^L}^{(t+1)} = \sum_{i \in C_k^L} w_i^{(t)} \theta_i^{(t+1)}, \quad (16)$$

$$\theta^{(t+1)} = \sum_{k \in \mathcal{K}_L} v_k^{(t)} \theta_{C_k^L}^{(t+1)}. \quad (17)$$

There exist constants  $C_\sigma, C_d, C_{\text{het}}, C_D > 0$ , such that, for  $\eta \leq c/(LE)$ ,

$$\frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E} \left[ \left\| \nabla \tilde{F}_L(\theta^{(t)}) \right\|^2 \right] \leq \frac{2(\tilde{F}_L(\theta^{(0)}) - \tilde{F}_L^\star)}{\eta ET} + C_\sigma \eta \sigma^2 + C_d L^2 \eta^2 E^2 G^2 + C_{\text{het}} (\epsilon_{\text{inter}}^2 + \epsilon_{\text{intra}}^2) + C_D B^2. \quad (18)$$

**Proof of Theorem 1.** By **Definition 3** and **Assumption 5**, the expected aggregation aligns with the time-averaged weights  $\bar{w}_i, \bar{v}_k$ . Using  $L$ -smoothness of  $\tilde{F}_L$  on the level set  $\mathcal{L}_0$ , according to **Defs. 1, 4** and **Assumption 1**, we have

$$\mathbb{E} \left[ \tilde{F}_L(\theta^{(t+1)}) \right] \leq \mathbb{E} \left[ \tilde{F}_L(\theta^{(t)}) \right] + \mathbb{E} \left[ \langle \nabla \tilde{F}_L(\theta^{(t)}), \Delta^{(t)} \rangle \right] + \frac{L}{2} \mathbb{E} [\|\Delta^{(t)}\|^2], \quad (19)$$

where  $\Delta^{(t)} := \theta^{(t+1)} - \theta^{(t)}$ . By local-SGD expansion on  $F_i^l$  in **Definition 2** using **Assumption 2** and **Assumption 6**,

$$\begin{aligned} \mathbb{E} \left[ \theta_i^{(t+1)} \right] &= \theta^{(t)} - \eta E \nabla F_i^l(\theta^{(t)}) + \mathbf{r}_i^{(t)}, \\ \mathbb{E} [\|\mathbf{r}_i^{(t)}\|^2] &\leq C_1 (\eta^2 E^2 L^2 G^2 + \eta^2 E \sigma^2). \end{aligned} \quad (20)$$

Taking the expectation over weights in **Assumption 5** yields

$$\mathbb{E} [\Delta^{(t)}] = -\eta E \sum_{k \in \mathcal{K}_L} \bar{v}_k \sum_{i \in C_k^L} \bar{w}_i \nabla F_i^l(\theta^{(t)}) + \mathbf{r}^{(t)}, \quad (21)$$

$$\mathbb{E} [\|\mathbf{r}^{(t)}\|^2] \leq C_2 (\eta^2 E^2 L^2 G^2 + \eta^2 E \sigma^2).$$

By **Defs. 3, 5** and **Assumption 3**,

$$\begin{aligned} \sum_{k \in \mathcal{K}_L} \bar{v}_k \sum_{i \in C_k^L} \bar{w}_i \nabla F_i^l(\theta) &= \nabla \tilde{F}_L(\theta) + \mathbf{b}(\theta), \\ \|\mathbf{b}(\theta)\|^2 &\leq C_3 (\epsilon_{\text{inter}}^2 + \epsilon_{\text{intra}}^2). \end{aligned} \quad (22)$$

Moreover, the distillation-induced structural error is bounded by **Assumption 4** as  $B^2$ . Substituting into **Eq. (19)**, and applying Young's inequality to handle cross terms, we obtain the one-round descent

$$\begin{aligned} \mathbb{E} \left[ \tilde{F}_L(\theta^{(t+1)}) \right] &\leq \mathbb{E} \left[ \tilde{F}_L(\theta^{(t)}) \right] - \frac{\eta E}{2} \mathbb{E} [\|\nabla \tilde{F}_L(\theta^{(t)})\|^2] \\ &\quad + C_\sigma \eta^2 E \sigma^2 + C_d L^2 \eta^2 E^2 G^2 + C_{\text{het}} (\epsilon_{\text{inter}}^2 + \epsilon_{\text{intra}}^2) + C_D B^2. \end{aligned} \quad (23)$$

Summing **Eq. (23)** from  $t = 0$  to  $T - 1$  and dividing by  $\frac{\eta ET}{2}$  gives **Eq. (18)**.  $\square$

**Table 4**

Comparison of datasets used in the experiments.

Dataset Name (Abbr.)	Dataset Size	Image Type	Classes	Characteristics
CIFAR10 (CF)	Training: 50000, Testing: 10000	RGB	10	Natural object
Fashion-MNIST (FM)	Training: 60000, Testing: 10000	Grayscale	10	Fashion product
Hydrophobicity of Composite Insulators (HY)	Training: 3626, Testing: 910	RGB	7	Insulator hydrophobicity in smart grid

## 5. System implementation and evaluation

### 5.1. Experimental setup

**Blockchain and Smart Contract Implementation:** The experimental evaluation is conducted on the consortium blockchain FISCO-BCOS v3.5.0. The network and evaluation configuration consisted of the following: 1) a blockchain with 15 consensus nodes deployed on the SA5.2XLARGE32 server (8-core CPU, 32GB RAM) and operating under the PBFT consensus mechanism, and 2) smart contracts implemented in Solidity v0.8.11, with client integration through Java SDK v3.5.0, and stress testing through Java SDK Demo v3.5.0. For automated vulnerability detection, we employed Echidna [33] fuzzer with echidna-parade for multicore fuzzing, which provides sophisticated configuration diversification.

**Datasets and Partitioning:** Two benchmark datasets CIFAR-10 (CF) [34] and Fashion-MNIST (FM) [35], and one smart grid dataset Hydrophobicity of Composite Insulators (HY) [36] are employed for comprehensive validation of general capability and practical utility in smart grid applications. Hydrophobicity, the water-repellent property of insulator surfaces, is a key indicator of insulation integrity, where degradation leads to increased leakage current, flashovers, and potential cascading grid failures, particularly in contaminated or humid environments. The HY dataset comprises data of composite insulators with varying hydrophobicity classifications (HC1-HC7), supporting critical power transmission infrastructure monitoring in smart grid systems. Detailed dataset characteristics are provided in [Table 4](#).

To simulate Non-IID nature of real-world federated learning, three partitioning strategies are employed: 1) Probabilistic Label Skew: Dirichlet distribution[37] with parameter ( $\alpha$ ) to create non-IID partitions where different clients have varying proportions of different classes, a smaller ( $\alpha$ ) leads to more heterogeneous distributions. 2) Shard-based Label Skew: A class-shard-based partitioning diversity distribution scheme[31] where each client is allocated samples from only a fraction ( $v$ ) of the total classes. 3) Quantity Skew: Data volume imbalance characterized by an imbalance factor ( $im$ ) combined with the Dirichlet distribution ( $\alpha$ ), simulating scenarios where clients have varying amounts of data.

[Fig. 4](#) demonstrates the comparative analysis of data distribution patterns across multiple partitioning strategies and presents sample visualizations from the three datasets, organized into four parts: [Fig. 4\(a–c\)](#), [\(d–f\)](#), and [\(g–i\)](#) show the data distributions for the CF, FM, and HY datasets, respectively. The distribution plots visualize the number of samples per client across different classes. Additionally, [Fig. 4\(j\)–\(l\)](#) provide representative sample images from the CF, FM, and HY datasets, respectively.

**Feature Extraction and Client Clustering:** Specialized autoencoders are implemented for the CF, FM, and HY datasets to facilitate feature extraction and clustering. For the FM dataset, the autoencoder comprises 4 encoding layers and 4 decoding layers, flattening the input images into 784-dimensional vectors and reducing the dimensionality to two dimensions. For the CF and HY datasets, autoencoders with convolutional layers perform dimensionality reduction through 3 convolutional and pooling layers to extract low-dimensional feature representations. The K-means algorithm was selected for clustering due to its computational efficiency, which is paramount for scalability in scenarios with a large number of clients.

**Table 5**

Clients training parameters configuration in FL tasks.

Parameter	Value
Optimizer	SGD
Learning rate	0.1
Batch size	32
Temperature	2
Loss	$\omega_C \cdot \mathcal{L}_C + \omega_K \cdot \mathcal{L}_K$
KDLoss weight $\omega_K$	0.3
CELoss weight $\omega_C$	0.7

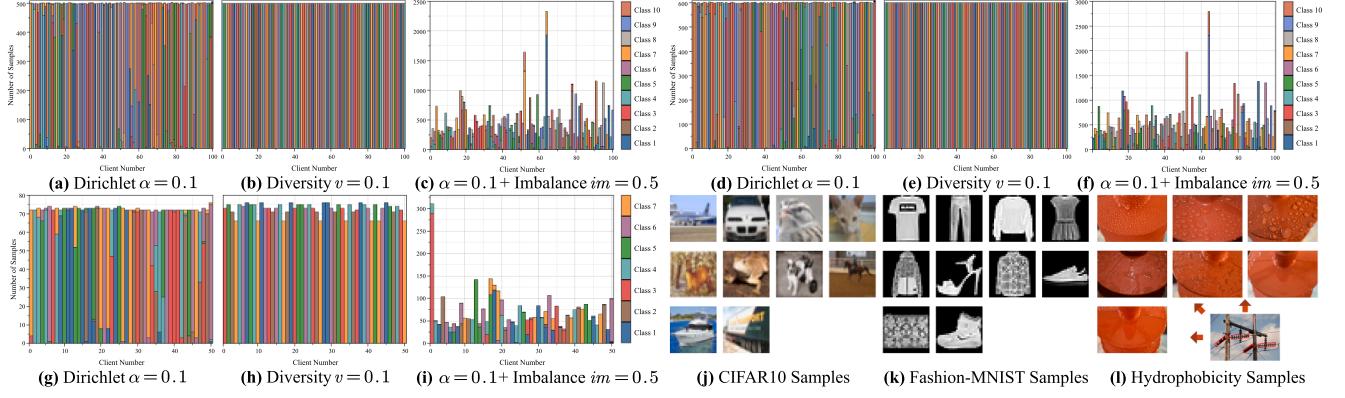
**Client Participation and Malicious Simulation:** To evaluate system robustness under dynamic conditions, we examine various client participation scenarios with participation ratios of 70%, 80%, 90%, and 100%, representing low, medium, high, and full device availability, respectively. These scenarios reflect realistic deployment conditions where client availability fluctuates due to network connectivity and device constraints. To assess robustness against adversarial attacks, we simulate label-flipping attacks, a prevalent and effective poisoning strategy in federated learning. We incorporate 20% malicious participants to create a realistic adversarial environment. The attack employs a deterministic data poisoning scheme where class labels are systematically altered to introduce conflicts in the training data, thereby undermining the formation of coherent decision boundaries in the global model.

**FL Models and Training Parameters:** The experiments involve training the MobileNetV3\_Small[38] model as the student model on clients with relatively low computational power. The federated learning framework is implemented using a PyTorch backend with three key modifications: 1) adjustment of the input layer to accommodate dataset-specific image dimensions, 2) reconfiguration of the classification layer to match the number of target categories, and 3) implementation of a layer-wise freezing strategy that maintains only the final classification layer as trainable to enhance computational efficiency. For knowledge distillation, a pre-trained ResNet34[39] serves as the teacher model, selected for its strong representation capabilities. The knowledge distillation process implements the DIST method[30]. Training hyperparameters are optimized through grid search on a separate validation set for both baseline methods and our proposed approach, with specific configurations detailed in [Table 5](#).

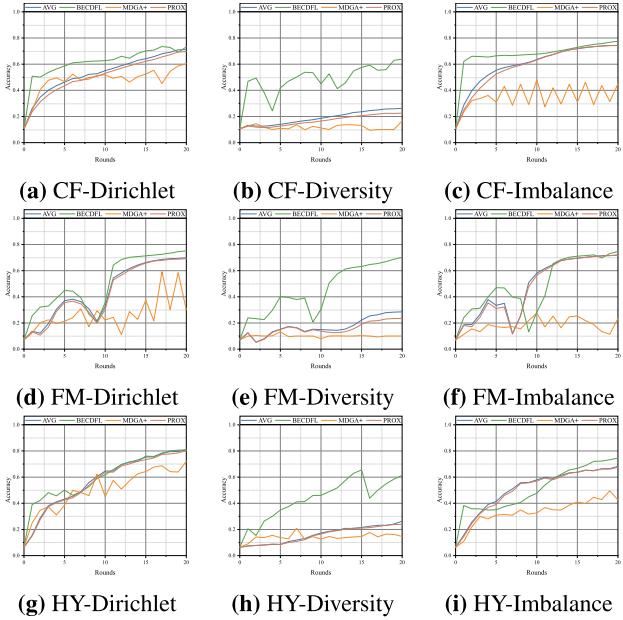
### 5.2. Experiments and results analysis

The experiments and analysis are conducted from three perspectives that align with the design objectives: effectiveness, robustness, and sustainability. First, the effectiveness evaluation compares the proposed scheme with baseline methods under heterogeneous distributions. Second, the robustness evaluation compares the stability with baseline methods under varying participation rates and in the presence of malicious behavior. Finally, the sustainability analysis investigates system overhead, with a particular focus on the blockchain and smart contracts.

**1) Effectiveness of BECDFL and Incentive Mechanism:** To comprehensively evaluate the performance of our proposed BECDFL approach, three representative and highly relevant federated learning methods (FedAvg [31], FedProx [32] and FedMDGA + [40]) are selected as baselines. FedAvg (AVG) is the canonical federated learning algorithm. It serves as the baseline to establish a performance reference point. By



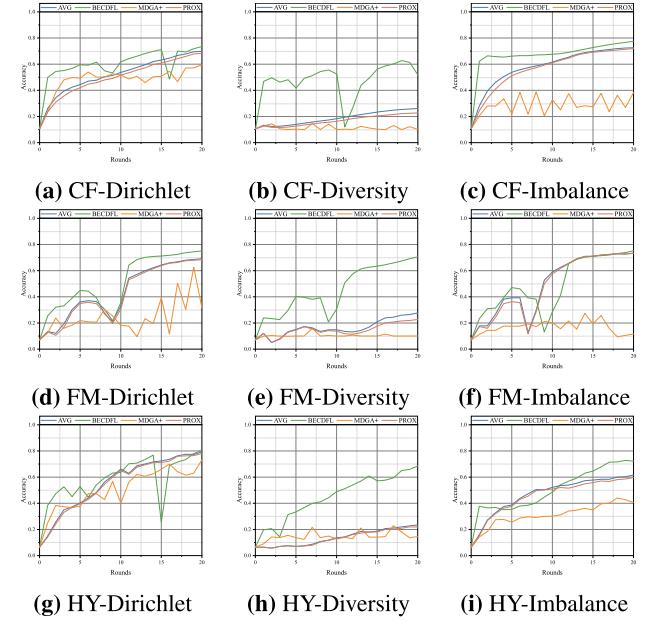
**Fig. 4.** Comparison of Dirichlet, diversity and imbalance distribution strategies with representative samples from CIFAR10, Fashion-MNIST and Hydrophobicity datasets.



**Fig. 5.** Performance analysis under full client participation across different distributions on CF, FM and HY datasets.

comparing against FedAvg, we can quantify the overall performance gain achieved by the advanced mechanisms in BECDFL and other baselines. FedProx (PROX) is selected as a foundational baseline for its approach to system heterogeneity at the local training stage. It modifies the local training objective with a proximal term that constrains local models to a neighborhood of the global model, thereby preventing model drift on devices with varying computational resources and enhancing convergence stability. FedMDGA + (MDGA +) is a sophisticated approach to data heterogeneity at the aggregation stage. It employs a constrained optimization framework to derive aggregation coefficients that minimize the L2-norm of the composite model update. This juxtaposition allows for a critical evaluation of a globally optimal mathematical solution versus a divide-and-conquer structural approach.

The effectiveness of BECDFL in achieving rapid convergence and high accuracy is demonstrated in Fig. 5, which compares the performance under full client participation. Compared to baseline methods, BECDFL consistently exhibits superior performance in both convergence speed and final accuracy. Under full client participation, BECDFL demonstrates effective initial learning, achieving 40-50% accuracy within the first 5 rounds across all datasets, outpacing baseline methods, which require 8-10 rounds to reach similar levels. This advantage



**Fig. 6.** Performance analysis under full client participation against malicious clients across different distributions on CF, FM and HY datasets.

is particularly evident with the Dirichlet distribution (Fig. 5(a),(d),(g)), where BECDFL reaches stable high performance 3-4 rounds faster than other baseline methods. This advantage is further highlighted by its performance with the diversity distribution (Fig. 5(b),(e),(h)), maintaining a consistent 25-35% performance advantage over baseline methods.

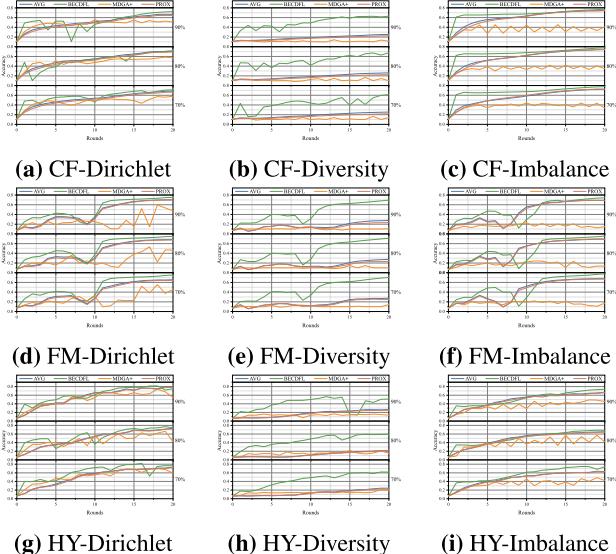
In the presence of 20% malicious clients (Fig. 6), the effectiveness of BECDFL extends to handling adversarial environments. The approach maintains effective learning with only minimal performance degradation, while baseline methods suffer severe effectiveness losses exceeding 20%. The proposed approach's effectiveness in rapid recovery from perturbations, typically within 2-3 rounds, further validates its practical utility in real-world scenarios.

**2) Robustness of BECDFL System Applicability:** The robustness of BECDFL is further validated through its resilience to varying client participation rates, heterogeneous distributions, and adversarial conditions, as demonstrated in Figs. 7 and 8. The performance variations across datasets can be attributed to their intrinsic properties. The FM dataset, with its simpler grayscale images, allows for more accurate feature extraction and clustering, resulting in higher accuracy. In contrast, the CF and HY datasets, with their complex color images and significant intra-class variation, pose a greater challenge. Under normal operating

**Table 6**

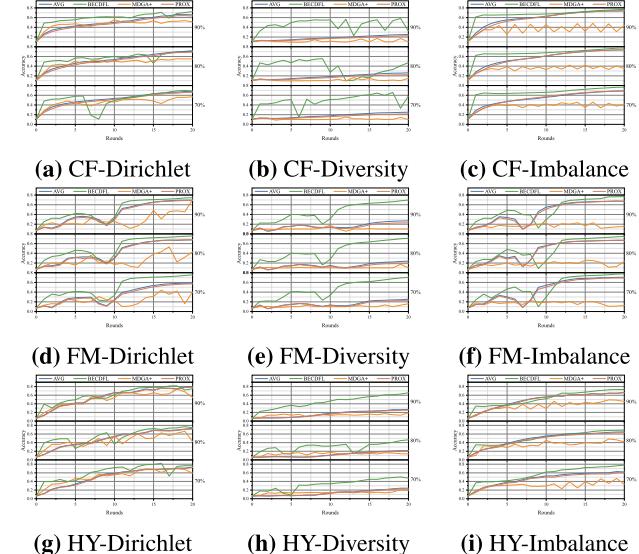
Mapping of algorithms to smart contract functions and gas consumption analysis.

Algorithm	Smart Contract	Function Name	Description	Function Notes
<b>Algorithm 1</b>	SC.1	maintainDeviceInfo	Add/update device info; verify existence and capability.	–
<b>Algorithm 2</b>	SC.2	uploadFeatures	Upload device features for clustering.	–
		submitClusterResults	Submit cluster results; assign device roles.	Scales with devices.
<b>Algorithm 3</b>	SC.3	submitCosineSimilarity/submitMarginalLoss	Submit model cosine similarity; Submit model marginal loss.	–
		updateClusterQualityScore	Calculate device quality score.	Scales with devices.
		updateClusterWeights	Compute aggregation weight in cluster.	Scales with clusters.
<b>Algorithm 4</b>	SC.4	submitClusterAggregation	Submit intra-cluster aggregation results.	Scales with clusters.
		submitGlobalAggregation	Submit inter-cluster aggregation results.	–

**Fig. 7.** Performance analysis under different participation rates (70%, 80%, 90%) across different distributions on CF, FM and HY datasets.

conditions (Fig. 7), BECDFL consistently maintains high accuracy compared to baseline methods across all tested participation rates and data distributions. Particularly noteworthy is its robustness under Diversity distributions (Fig. 7(b),(e),(h)) where it achieves 30-40% higher accuracy than competing approaches. For mission-critical smart grid applications, the HY dataset exemplifies this stability. On the HY dataset with Imbalance distribution and 70% participation, BECDFL maintains stable accuracy 30-40% higher than competing approaches with less fluctuation, as shown in Fig. 7(h), which is crucial for accurate insulator condition monitoring in power systems.

The system's robustness against adversarial attacks, a concern for smart grid security, is further evidenced in Fig. 8. Even with 20% of clients injecting malicious updates, BECDFL effectively mitigates accuracy degradation across all scenarios, whereas baseline methods suffer significant performance degradation. In the CF dataset with Diversity distribution (Fig. 8(b)), while baseline methods suffer severe degradation, dropping to 15-20% accuracy), BECDFL maintains an accuracy of 45-50%, representing only a 15% decrease from its non-malicious performance. Similarly, in the FM dataset, BECDFL maintains stable performance around 55-60% accuracy with 90% participation (Fig. 8(e)), despite malicious interference, outperforming the next best baseline by approximately 25%. Most critically for power system applications, in the Hydrophobicity dataset analysis (Fig. 8(h)), BECDFL's performance only degrades by 10-12% under malicious conditions, maintaining an accuracy of 55-58% compared to baseline methods which drop below 35%. This resilience across different participation rates shows minimal variance in performance degradation, demonstrating BECDFL's robust applicability in real-world smart grid deployments where system security and reliability are paramount.

**Fig. 8.** Performance analysis under different participation rates (70%, 80%, 90%) against malicious clients across different distributions on CF, FM and HY datasets.

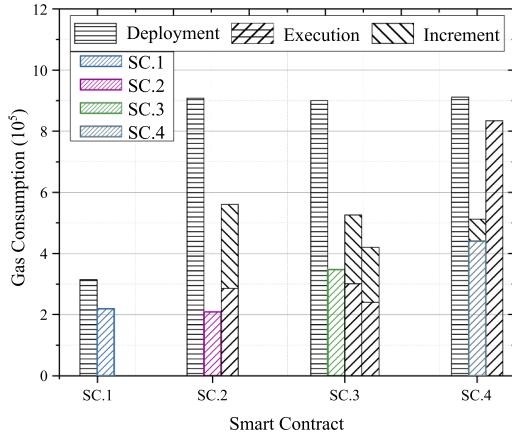
**3) Sustainability of the Blockchain-Empowered System:** The sustainability of the BECDFL system is assessed through an analysis of the gas efficiency and scalability of its underlying smart contracts, encompassing gas consumption, transaction throughput, and latency distribution. Core algorithms are mapped to their specific smart contract implementations, detailing the functions and associated costs for each component, as presented in Table 6, which serves as the foundation for subsequent performance evaluation. Notably, the Function Notes column in Table 6 identifies which functions have execution costs that scale with the increasing number of devices or clusters, which is critical for system scalability assessment.

Fig. 9(a) visually illustrates the gas overhead, providing critical insights into the system's operational efficiency, the specific gas consumption values are detailed in Table 7. The Fig. 9(a) displays three key metrics for each contract: deployment cost (initial cost), execution cost (operational cost), and increment cost (additional marginal consumption). Table 7 quantifies the computational cost of each function in Gas units. Crucially, unlike in Ethereum, in a consortium blockchain, Gas serves not as a direct monetary cost but as a standardized metric for resource consumption. A lower Gas cost signifies higher computational efficiency.

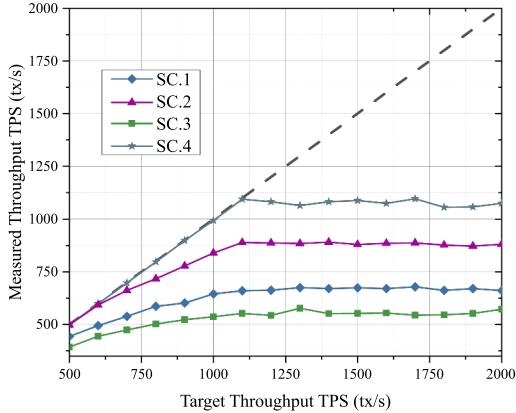
The data reveals that SC.1 is the most lightweight contract, while SC.4's global aggregation function is the most resource-intensive, which aligns with the functional complexity outlined in Table 6. Smart Contract 4 (SC.4) exhibits a deployment cost of approximately  $9 \times 10^5$  Gas, an execution cost of approximately  $4.4 \times 10^5$  Gas, and a significantly lower increment cost of just under  $1 \times 10^5$  Gas. This underscores the gas-efficient contract design, where the incremental cost for adding new devices to functions is minimized. This low incremental gas consumption is critical for the scalability of device participation, ensuring that on-

**Table 7**  
Detailed gas consumption for smart contract deployment and execution (in Gas Units).

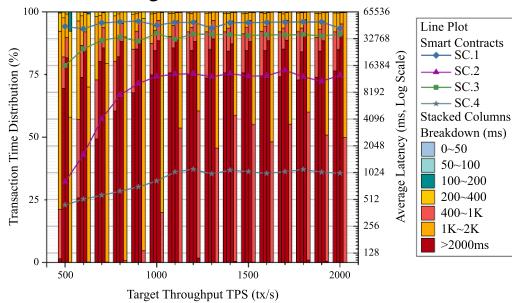
Smart Contract	Deployment Cost	Execution Function Cost (Incremental Cost)		
		Function 1 (Increment)	Function 2 (Increment)	Function 3 (Increment)
SC. 1 (Algo. 1)	31449	21893	-	-
SC. 2 (Algo. 2)	90838	20880	28608 (27478)	-
SC. 3 (Algo. 3)	89989	34756	30142 (22439)	24001 (17999)
SC. 4 (Algo. 4)	91139	44018 (7237)	83439	-



(a) Overhead of Deployment and Execution of Algorithms Implemented through Smart Contracts



(b) Performance Analysis of Selected Smart Contract Functions in High-Load Scenarios



(c) Transaction Latency Distribution Analysis under Variable Load Conditions

**Fig. 9.** Gas-efficient and scalable smart contracts: a performance analysis for sustainability of blockchain-empowered system.

boarding a large number of users does not lead to prohibitive gas fees. Similarly, Smart Contract 3 (SC.3), implementing the Quality Assessment algorithm, shows comparable deployment costs to SC.4 but with lower execution and increment costs, of approximately  $3.5 \times 10^5$  and  $4.2 \times 10^5$  Gas, respectively. This quantitative data is vital for predicting

the system's load on the consortium's shared infrastructure. Notably, the colored bars in Fig. 9(a) represent specific functions within these smart contracts that are further analyzed for performance under high-load scenarios, as shown in Fig. 9(b) and (c). Throughput, as shown in Fig. 9(b), measures the aggregate capacity of the entire pipeline. Latency, visualized in Fig. 9(c), measures the time taken for a single transaction to complete its entire journey.

Fig. 9(b) plots measured throughput transactions per second (TPS) against target throughput TPS, with the diagonal dashed line representing ideal performance. SC.4 demonstrates superior performance with sustained throughput of approximately 1100 tx/s even under high load conditions, while SC.1 maintains around 850 tx/s. SC.2 and SC.3 show lower but stable throughput at approximately 650 tx/s and 550 tx/s, respectively. Each block contains multiple transactions that the consensus mechanism validates as a single batch. Therefore, the TPS of approximately 1100 tx/s for SC.4 does not mean each transaction takes less than 1ms. This indicates that the system achieves block finalization containing a substantial number of transactions, with finalized transactions emerging at an average throughput of 1,100 transactions per second. The system demonstrates high-volume processing capability, maintaining consistent throughput even with extended individual transaction processing latencies. This sustained performance under pressure indicates the system's capacity to handle frequent model updates and accommodate a large influx of participating devices.

Fig. 9(c) presents a detailed analysis of the blockchain system's latency characteristics under progressively increasing loads. It is essential to note that the stacked columns, which illustrate the percentage-based breakdown of transaction latency, correspond to the left Y-axis. Concurrently, the four line plots, which track the average transaction finality latency for each of the four smart contracts, are mapped to the logarithmic right Y-axis. The stacked columns provide a perspective on the system's response time consistency. At lower loads, a portion of transactions are processed in the faster latency buckets. As the system is pushed towards its peak capacity, there is a discernible shift in the distribution towards longer latencies. Complementing this distributional analysis, the line plots reveal that the average latencies for SC.1 and SC.3 are shown to be relatively high, suggesting that these operations are inherently computationally or I/O intensive. In contrast, the latencies for SC.2 and SC.4 exhibit a dynamic relationship with system load, beginning at lower average values and exhibiting an upward trend that eventually stabilizes as the system reaches its maximum throughput. This comprehensive view validates the system's soundness and its suitability for deployment in demanding environments.

## 6. Conclusion

This paper proposes a BECDFL framework for smart grid environments that effectively addresses both heterogeneity and scalability challenges. The framework establishes a secure infrastructure through blockchain integration, ensuring data integrity and system trust throughout the federated learning process. By implementing computing power categorization and data distribution clustering, the proposed approach systematically addresses multifaceted heterogeneity challenges. A quality-aware cluster aggregation mechanism based on knowledge distillation enables computationally constrained devices to effectively participate in model training by learning from more powerful models. This mechanism evaluates both model and data quality through

cosine similarity and marginal loss metrics to enhance result reliability in large-scale deployments. The evaluation results confirm performance improvements in model accuracy, convergence efficiency, and resilience against data heterogeneity in complex smart grid contexts. Specifically, the proposed framework achieves stable performance 3-4 rounds faster than baselines and maintains a higher accuracy under severe data heterogeneity and fluctuating client participation. Critically, the framework demonstrates significant robustness, exhibiting only a 10-12% accuracy drop against 20% malicious clients, a scenario under which baselines collapse. In future work, we will explore dynamic adjustment mechanisms for cluster formation and incentive parameters to enhance adaptability to evolving smart grid conditions, and investigate the integration of explainable AI techniques for critical power system applications.

### CRediT authorship contribution statement

**Zhihao Zhou:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization; **Yunhua He:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization; **Wei Zhang:** Writing – original draft, Validation, Software, Project administration, Investigation, Formal analysis, Conceptualization; **Zhaohui Ding:** Writing – review & editing, Visualization, Supervision, Resources, Methodology, Funding acquisition, Data curation; **Bin Wu:** Writing – original draft, Supervision, Project administration, Funding acquisition, Conceptualization; **Ke Xiao:** Writing – review & editing, Supervision, Project administration, Investigation, Conceptualization.

### Data availability

The data that has been used is confidential.

### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Acknowledgments

This work was supported by the [National Natural Science Foundation of China \(62272007, U23B2002\)](#), in part by [Beijing Municipal Natural Science Foundation](#) under Grant [L222002](#), and in part by the Excellent Young Talents Project of the Beijing Municipal University Teacher Team Construction Support Plan under Grant [BPHR202203031](#).

### References

- [1] X. Fang, S. Misra, G. Xue, D. Yang, Smart Grid — the New and Improved Power Grid: A Survey, *IEEE Commun. Surv. Tut.* 14 (4) (2012) 944–980. <https://doi.org/10.1109/SURV.2011.101911.00087>
- [2] P. Thunshirin, I. Mlinaric, J. Berg, A Qualitative Analysis of Consumer Motivations and Barriers towards Active Smart Meter Utilization, *Energy Policy* 203 (2025) 114623. <https://doi.org/10.1016/j.enpol.2025.114623>
- [3] P. Wilczek, Connecting the dots: distribution grid investments to power the energy transition, *IET Conf. Proc.* 2021 (2022) 1–18. <https://digital-library.theiet.org/doi/abs/10.1049/icp.2021.2473>. <https://doi.org/10.1049/icp.2021.2473>
- [4] B.P. Bhattachari, S. Paudyal, Y. Luo, M. Mohanpurkar, K. Cheung, R. Tonkoski, R. Hovsepian, K.S. Myers, R. Zhang, P. Zhao, M. Manic, S. Zhang, X. Zhang, Big Data Analytics in Smart Grids: State-of-the-art, Challenges, Opportunities, and Future Directions, *IET Smart Grid* 2 (2) (2019) 141–154. <https://doi.org/10.1049/iet-stg.2018.0261>
- [5] M.M. Alam, M.J. Hossain, M.A. Habib, M.Y. Arifat, M.A. Hannan, Artificial Intelligence Integrated Grid Systems: Technologies, Potential Frameworks, Challenges, and Research Directions, *Renewable and Sustainable Energy Reviews* 211 (2025) 115251. <https://doi.org/10.1016/j.rser.2024.115251>
- [6] G.S. Wagh, S. Mishra, A. Agarwal, A Distributed Privacy-Preserving Framework with Integrity Verification Capabilities for Metering Data and Dynamic Billing in a Malicious Setting, *IEEE Transactions on Smart Grid* 15 (6) (2024) 5813–5825. <https://doi.org/10.1109/TSG.2024.3433385>
- [7] J. Guo, Z. Liu, S. Tian, F. Huang, J. Li, X. Li, K.K. Igorevich, J. Ma, TFL-DT: A Trust Evaluation Scheme for Federated Learning in Digital Twin for Mobile Networks, *IEEE J. Selected Areas Commun.* 41 (11) (2023) 3548–3560. <https://doi.org/10.1109/JSCA.2023.3310094>
- [8] L. Yang, Y. Miao, Z. Liu, Z. Liu, X. Li, D. Kuang, H. Li, R.H. Deng, Enhanced Model Poisoning Attack and Multi-Strategy Defense in Federated Learning, *IEEE Trans. Inf. Forensics Secur.* 20 (2025) 3877–3892. <https://doi.org/10.1109/TIFS.2025.3555193>
- [9] H.-Y. Tran, J. Hu, X. Yin, H.R. Pota, An Efficient Privacy-Enhancing Cross-Silo Federated Learning and Applications for False Data Injection Attack Detection in Smart Grids, *IEEE Transactions on Information Forensics and Security* 18 (2023) 2538–2552. <https://doi.org/10.1109/TIFS.2023.3267892>
- [10] Q. Mao, L. Wang, Y. Long, L. Han, Z. Wang, K. Chen, A Blockchain-Based Framework for Federated Learning with Privacy Preservation in Power Load Forecasting, *Knowl. Based Syst.* 284 (2024) 111338. <https://doi.org/10.1016/j.knosys.2023.111338>
- [11] H. Zhu, J. Xu, S. Liu, Y. Jin, Federated Learning on Non-IID Data: A Survey, *Neurocomputing* 465 (2021) 371–390. <https://doi.org/10.1016/j.neucom.2021.07.098>
- [12] Y. Fathy, P. Barnaghi, R. Tafazolli, Large-Scale Indexing, Discovery, and Ranking for the Internet of Things (IoT), *ACM Comput. Surv.* 51 (2) (2018) 29:1–29:53. <https://doi.org/10.1145/3154525>
- [13] A. Shobol, M.H. Ali, M. Wadi, M.R. Tür, Overview of Big Data in Smart Grid, in: 2019 8th International Conference on Renewable Energy Research and Applications (ICRERA), 2019, pp. 1022–1025. <https://doi.org/10.1109/ICRERA47325.2019.8996527>
- [14] H. Wen, X. Liu, B. Lei, M. Yang, X. Cheng, Z. Chen, A Privacy-Preserving Heterogeneous Federated Learning Framework with Class Imbalance Learning for Electricity Theft Detection, *Applied Energy* 378 (2025) 124789. <https://doi.org/10.1016/j.apenergy.2024.124789>
- [15] C. Fan, R. Chen, J. Mo, L. Liao, Personalized Federated Learning for Cross-Building Energy Knowledge Sharing: Cost-effective Strategies and Model Architectures, *Applied Energy* 362 (2024) 123016. <https://doi.org/10.1016/j.apenergy.2024.123016>
- [16] C. Lu, S. Deng, Y. Wu, H. Zhou, W. Ma, Federated Learning Based on OPTICS Clustering Optimization, *Discrete Dynam. Nature Soc.* 2022 (2022) e7151373. <https://doi.org/10.1155/2022/7151373>
- [17] J. Lee, H. Ko, Energy and Distribution-Aware Cooperative Clustering Algorithm in Internet of Things (IoT)-Based Federated Learning, *IEEE Trans. Veh. Technol.* 72 (10) (2023) 13799–13804. <https://doi.org/10.1109/TVT.2023.3277438>
- [18] Z. Wang, P. Yu, H. Zhang, Privacy-Preserving Regulation Capacity Evaluation for HVAC Systems in Heterogeneous Buildings Based on Federated Learning and Transfer Learning, *IEEE Trans. Smart Grid* 14 (5) (2023) 3535–3549. <https://doi.org/10.1109/TSG.2022.3231592>
- [19] C. Tong, L. Zhang, Y. Ding, D. Yue, Mutual Knowledge-Distillation-Based Federated Learning for Short-Term Forecasting in Electric IoT Systems, *IEEE Internet of Things J.* 11 (19) (2024) 31190–31205. <https://doi.org/10.1109/JIOT.2024.3416527>
- [20] F. Xiao, L. Wu, Personalized Federated Lithium-Ion Battery Capacity Prediction via Cluster and Fusion Modules, *IEEE Trans. Transp. Electrific.* 10 (3) (2024) 6434–6448. <https://doi.org/10.1109/TTE.2023.3334826>
- [21] H.U. Manzoor, A. Jafri, A. Zoha, Adaptive Single-Layer Aggregation Framework for Energy-Efficient and Privacy-Preserving Load Forecasting in Heterogeneous Federated Smart Grids, *Internet of Things* 28 (2024) 101376. <https://doi.org/10.1016/j.iot.2024.101376>
- [22] R. Zhao, J. Lu, Z. Liu, T. Wang, W. Guo, T. Lan, C. Hu, A Federated-Learning Algorithm Based on Client Sampling and Gradient Projection for the Smart Grid, *Electronics* 13 (11) (2024) 2023. <https://doi.org/10.3390/electronics13112033>
- [23] H. Wu, Z. Xu, Multi-Energy Load Forecasting in Integrated Energy Systems: A Spatial-Temporal Adaptive Personalized Learning Approach, *IEEE Trans. Ind. Inf.* 20 (10) (2024) 12262–12274. <https://doi.org/10.1109/TII.2024.3417297>
- [24] N. Abdulla, M. Demirci, S. Ozdemir, Smart Meter-Based Energy Consumption Forecasting for Smart Cities Using Adaptive Federated Learning, Sustainable Energy, Grids and Networks 38 (2024) 101342. <https://doi.org/10.1016/j.segan.2024.101342>
- [25] Y. He, Y. Yang, C. Wang, A. Xie, L. Ma, B. Wu, Y. Wu, A Verifiable EVM-based Cross-Language Smart Contract Implementation Scheme for Matrix Calculation, *Digital Commun. Netw.* 11 (2) (2025) 432–441. <https://doi.org/10.1016/j.dcan.2024.03.003>
- [26] Y. He, M. Luo, B. Wu, L. Sun, Y. Wu, Z. Liu, K. Xiao, A Game Theory-Based Incentive Mechanism for Collaborative Security of Federated Learning in Energy Blockchain Environment, *IEEE Internet of Things J.* 10 (24) (2023) 21294–21308. <https://doi.org/10.1109/JIOT.2023.3282732>
- [27] S. Rajasekaran, Efficient parallel hierarchical clustering algorithms, *IEEE Trans. Parallel Distrib. Syst.* 16 (6) (2005) 497–502. <https://doi.org/10.1109/TPDS.2005.72>
- [28] D. Deng, DBSCAN Clustering Algorithm Based on Density, in: 2020 7th International Forum on Electrical Engineering and Automation (IFEEA), 2020, pp. 949–953. <https://doi.org/10.1109/IFEEA51475.2020.00199>
- [29] Y. Zhao, A.K. Srivastava, K.L. Tsui, Regularized Gaussian Mixture Model for High-Dimensional Clustering, *IEEE Trans. Cybern.* 49 (10) (2019) 3677–3688. <https://doi.org/10.1109/TCYB.2018.2846404>
- [30] T. Huang, S. You, F. Wang, C. Qian, C. Xu, Knowledge Distillation from a Stronger Teacher, in: Proceedings of the 36th International Conference on Neural Information Processing Systems, NIPS '22, Curran Associates Inc., Red Hook, NY, USA, 2024, pp. 33716–33727.

- [31] B. McMahan, E. Moore, D. Ramage, S. Hampson, B.A. y Arcas, Communication-Efficient Learning of Deep Networks from Decentralized Data, in: Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, PMLR, 2017, pp. 1273–1282.
- [32] T. Li, A.K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, V. Smith, Federated Optimization in Heterogeneous Networks, in: I. Dhillon, D. Papailiopoulos, V. Sze (Eds.), Proceedings of Machine Learning and Systems, 2, 2020, pp. 429–450. [https://proceedings.mlsys.org/paper\\_files/paper/2020/file/1f5fe83998a09396ebe6477d9475ba0c-Paper.pdf](https://proceedings.mlsys.org/paper_files/paper/2020/file/1f5fe83998a09396ebe6477d9475ba0c-Paper.pdf).
- [33] A. Groce, G. Grieco, echidna-parade: a tool for diverse multicore smart contract fuzzing, in: Proceedings of the 30th ACM SIGSOFT International Symposium on Software Testing and Analysis, ISSTA 2021, Association for Computing Machinery, New York, NY, USA, 2021, p. 658–661. <https://doi.org/10.1145/3460319.3469076>
- [34] A. Krizhevsky, Learning Multiple Layers of Features from Tiny Images, 2009.
- [35] H. Xiao, K. Rasul, R. Vollgraf, Fashion-MNIST: a Novel Image Dataset for Benchmarking Machine Learning Algorithms, 2017. [arXiv:cs.LG/1708.07747](https://arxiv.org/abs/1708.07747)
- [36] C.-C.A. Kokalis, T. Tasakos, V.T. Kontargyri, G. Siolas, I.F. Gonos, Hydrophobicity Classification of Composite Insulators Based on Convolutional Neural Networks, Engineering Applications of Artificial Intelligence 91 (2020) 103613. <https://doi.org/10.1016/j.engappai.2020.103613>
- [37] Q. Li, Y. Diao, Q. Chen, B. He, Federated Learning on Non-IID Data Silos: An Experimental Study, in: 2022 IEEE 38th International Conference on Data Engineering (ICDE), 2022, pp. 965–978. <https://doi.org/10.1109/ICDE53745.2022.00077>
- [38] A. Howard, M. Sandler, B. Chen, W. Wang, L.-C. Chen, M. Tan, G. Chu, V. Vasudevan, Y. Zhu, R. Pang, H. Adam, Q. Le, Searching for MobileNetV3, in: 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019, pp. 1314–1324. <https://doi.org/10.1109/ICCV.2019.00140>
- [39] K. He, X. Zhang, S. Ren, J. Sun, Deep Residual Learning for Image Recognition, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (Cvpr), 2016, pp. 770–778. <https://doi.org/10.1109/CVPR.2016.90>
- [40] Z. Hu, K. Shaloudegi, G. Zhang, Y. Yu, Federated Learning Meets Multi-Objective Optimization, IEEE Trans. Netw. Sci. Eng. 9 (4) (2022) 2039–2051. <https://doi.org/10.1109/TNSE.2022.3169117>