IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2024.DOI

Adding Data Quality to Federated Learning Performance Improvement

ERNESTO G. VALENTE NETO [1], SOLON A. PEIXOTO JR. [2], VALDERI R. Q. LEITHARDT (Senior Member, IEEE), [3], JUAN F. P. SANTANA [4], JULIO C. S. DOS ANJOS [2,5,6]

[1]Federal University of Ceara, PPGETI, Fortaleza, PC 60020-181, Brasil (e-mail: gurgelvalente@alu.ufc.br)
[2]Federal University of Ceara Campus Itapaje, Itapaje, PC 62600-000, Brasil (e-mail: solon.alves@ufc.br)
[3]Instituto Universitário de Lisboa (ISCTE-IUL), ISTAR, Avenida das Forças Armadas, 1649-026, Lisboa, Portugal (e-mail: valderi.leithardt@iscte-iul.pt)
[4]University of Salamanca, Expert Systems and Applications Laboratory, Salamanca, Spain (e-mail: fcofds@usal.es)
[5]Federal University of Ceara, Graduate Program in Teleinformatics Engineering (PPGETI/UFC), Center of Technology, Campus of Pici, Fortaleza, Ceara, Brazil, 60455-970, (e-mail:jcsanjos@ufc.br)
[6]National Institute of Science and Technology (INCT-Signals), Campus of Pici, Fortaleza, Ceara, Brazil, 60455-970

Corresponding author: Ernesto G. Valente Neto (e-mail: gurgelvalente@alu.ufc.br).

This work was partially funded by Coordination for the Improvement of Higher Education Personnel - Brazil (CAPES) - Finance Code 001. A realização desta investigação foi parcialmente financiada profundos nacionais através da FCT - Fundação para a Ciência e Tecnologia, I.P. no âmbito dos projetos UIDB/04466/2020 e UIDP/04466/2020. This study was partially support of the National Institute of Science and Technology (INCT-Signals), sponsored by Brazil's National Council for Scientific and Technological Development (CNPq) under grant no. 406517/2022-3. Partial support was provided via the CEREIA Project (# 2020/09706-7) São Paulo Research Foundation (FAPESP), FAPESP–MCTIC-CGI.BR in partnership with the Hapvida NotreDame Intermedica group.

ABSTRACT Massive data generation from Internet of Things (IoT) devices increases the demand for efficient data analysis to extract meaningful insights. Federated Learning (FL) allows IoT devices to collaborate in AI training models while preserving data privacy. However, selecting high-quality data for training remains a critical challenge in FL environments with non-independent and identically distributed (non-iid) data. Poor-quality data introduces errors, delays convergence, and increases computational costs. This study develops a data quality analysis algorithm for FL and centralized environments to address these challenges. The proposed algorithm reduces computational costs, eliminates unnecessary data processing, and accelerates AI model convergence. The experiments used the MNIST, Fashion-MNIST, CIFAR-10, and CIFAR-100 datasets, and performance evaluation was based on main literature metrics like accuracy, recall, F1 score, and precision. Results show the best case execution time reductions of up to 56.49%, with an accuracy loss of around 0.50%.

INDEX TERMS Data Quality, Deep Learning, Federated Learning, IoT, IID, non-IID.

I. INTRODUCTION

Advances in Artificial Intelligence (AI) and Internet of Things (IoT) development significantly impact various sectors, enabling intelligent data-driven applications. However, this development brings challenges for data security and mobility [1], especially in the health sector, where these technologies play a key role in protecting patient privacy and ensuring compliance with confidentiality regulations [2].

In scenarios where data protection is paramount, approaches to effective monitoring and management stand out, ensuring regulatory compliance and the safety of patient data [3], [4]. In this context, the ability to process large amounts of data efficiently and collaboration between neural networks are becoming essential for various applications [5], [6].

From this perspective, Federated Learning (FL) emerges as a promising approach, allowing IoT devices to collaborate to create a neural network without sharing raw data. The edge devices share their local model variables by updating a global model, which updates the local devices until a global minimum is reached [7].

Despite its promise, this approach introduces several challenges, such as handling Independent and Identically Distributed (iid) and Non-Independent and Identically Distributed (non-iid) datasets [8]–[10]. This approach includes mitigating the high communication costs due to data dispersion and the complexity of selecting clients during model training. These factors affect learning convergence and prolong execution time, resulting in a more computationally intensive process on edge devices with hardware and energy

limitations.

Furthermore, establishing incentives for collaboration and managing device heterogeneity remain significant challenges [11], [12]. Consequently, addressing computational demands requires automating cost management, increasing resource availability [13], capturing temporal dependencies [14], and ensuring data quality at the edge [15].

In this context, state-of-the-art studies have explored solutions that underscore the significance of signal processing [16], focusing on device heterogeneity and data privacy protection. These approaches aim to optimize resource efficiency and improve the training of FL models while ensuring robust privacy support [11], [17].

Ultimately, these studies highlight key issues in managing large data volumes, high communication costs, and the complexity of data selection. Furthermore, the nature of non-iid data adds a layer of complexity to the training models.

Current FL approaches do not evaluate data quality at the edge, leading to low-quality inputs and inefficient training. Addressing this limitation, our model introduces an entropy-based data selection method to optimize FL performance with the **following:**

i. providing a data quality analysis algorithm on edge to select data with the best information with class and accuracy maintenance from the FL original models;
ii. reducing unnecessary data processing with low data quality to save energy on the edge; and
iii. improving FL the computation performance by the reducing execution time by 50% in IoT devices.

In general, at the top layer, the **FL Model and Aggregation Server** orchestrate clients (nodes), detect rare events, and ensure resistance to poisoning attacks or failures, such as communication cost [18]. The responsibilities may include techniques such as feature extraction [19], dynamic regularization [20], node selection [21], client clustering [22], client sampling [23], client contributions [24], and adaptive selection [25], as well as layers of security, increased fairness in collaboration between clients, and poisoning attack mitigation and defense mechanisms [26], [27].

For instance, **upper-layer aspects** — such as client orchestration and connection problem management — are commonly addressed in studies of global aggregation algorithms at the aggregation server in FL approaches. Alternatively, designers may integrate these aspects with broader solutions (e.g., cryptography, blockchain, and connection management)

In addition, the taxonomy in the Continuum of the Internet of Things [28], [29] assigns these responsibilities to the orchestrator, including how to manage connectivity, network resources, resource management, network management, and security across distributed edge, fog, and cloud layers. Some responsibilities are shared between FL algorithms and the orchestrator, especially in scenarios where coordination, data availability, and system resilience are crucial for distributed training and aggregation processes.

This work proposes an **agnostic algorithm** (i.e., a new AI layer as a data preparation step in the edge) **without interfering** with the **FL** algorithm execution **or the server** where aggregation occurs. The algorithm is based on edge data quality evaluation, which removes data without relevant information for training and improves the convergence of training algorithms by selecting the most significant information from the input data based on the entropy metric.

The **algorithm operates at the edge** and selects the best aggregated information for processing in the edge devices. This design choice ensures that the proposed approach does not rely on or intervene with the functionalities of the **FL Model or Aggregation Server**. As a result, the algorithm can be integrated into different FL pipelines regardless of the server's adopted aggregation strategies or management. Due to this, rare event detection, FL security, attacks that compromise the integrity or reliability of the neural model hosted on the aggregation server, and communication issues **are outside the scope of this work.**

The remainder of this paper is structured as follows. Section II provides an appropriate background and covers Related Work. Section IV shows the proposed model in detail. Section V presents the methodology and the evaluated scenarios. Section VI presents the evaluations and the results achieved. Section VIII presents the final considerations, Section IX outlines future directions, and finally, Section Appendix X contains more detailed information about the state-of-the-art.

II. PRELIMINARIES AND BACKGROUND

This section discusses three concepts: 1) Iid data with centralized learning, where independent and identically distributed data are collected centrally. 2) Non-iid data with FL addresses decentralized data, reflecting uneven distribution and non-independent data. 3) Entropy measures uncertainty or disorder. Finally, it presents the Related Work and discusses the main problems.

A. IID AND NON-IID DATA

The iid data follows the same probability distribution by observations in which the samples share the same probabilistic distribution, independent of their occurrence. For example, we assume that the data come from a production line source, which shows the same probability distribution. Therefore, inferences can be made regarding these characteristics.

The iid property assumes that each observation is independent of the others [30], meaning that one sample does not influence another. Such an assumption allows statistical tools, such as the law of large numbers and the central limit theorem, to generalize the results. However, these assumptions are often not satisfied in the real world. Data are commonly correlated or exhibit sample heterogeneity [31].

These conditions violate the iid data assumptions, where the observations do not follow the same probabilistic distribution, and the observations may be very close to each

other or strongly correlated rather than having more distant relationships. Furthermore, dynamic environments may make the iid assumption unrealistic.

The distributions change over time, leading to an effect known as concept drift, which represents a shift or evolution in the data that invalidates the created AI model. Figure 1 presents a visual representation of these concepts.

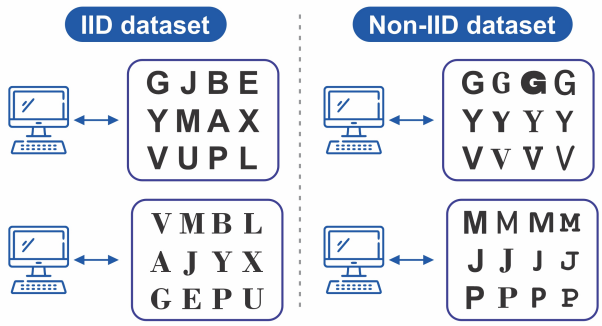


FIGURE 1: The iid and non-iid data.

Moreover, non-iid data violate at least one of the conditions that define iid data; each observation or sample must be independent of the others, and all samples originate from the same probabilistic distribution.

In other words, non-iid data involve correlated distributions, where samples may influence one another, and probabilistic distributions may vary among data. Additionally, non-iid datasets may present different subsets of data that follow different distributions.

B. CENTRALIZED LEARNING AND FEDERATED LEARNING

In centralized Machine Learning (ML), data originates and is stored, analyzed, and processed on a dedicated server or in a centralized location. This architecture promotes efficiency in statistical modeling and pattern detection. This model facilitates the application of ML algorithms that require large amounts of data to generalize well and achieve accurate predictive results.

However, centralized models present challenges, mainly related to data security and collecting and centralizing large volumes of data. Therefore, this exposes sensitive information to risks such as data leakage or cyberattacks [32]. Centralization also leads to issues related to latency, where data from different sources is centralized on a server, consuming a significant amount of communication, especially with geographically distributed data [33].

Unlike the centralized approach, FL follows a decentralized strategy for training ML models. Data from different sources contribute to training multiple devices or nodes in a network (clients), such as smartphones, tablets, IoT sensors, and other edge-computing devices. Each device uses its data to train an AI model and then sends model parameter updates to a central server that aggregates the updates from the parameters of the neural network. In this approach, private data remain on edge devices, never shared directly, respecting ethical and legal perspectives in contexts where data are sensitive. Furthermore, it reduces massive data transfers and the risk of large-scale data leaks [34].

Figure 2 compares ML architectures. In Figure 2. (1), the data from multiple devices are centralized and stored for model training. In contrast, Figure 2. (2) depicts decentralized training, where the data remains on the devices, and the model trains locally. A global aggregation algorithm combines updates from the neural networks of each device.

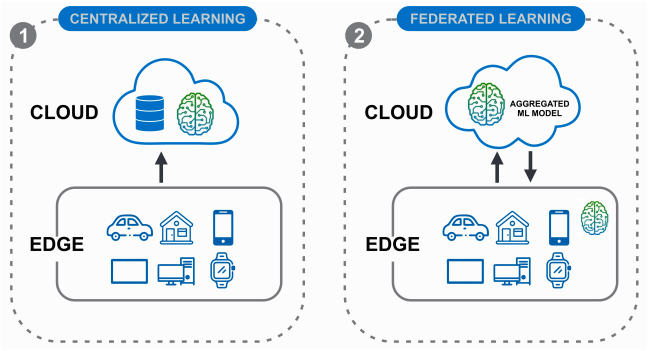


FIGURE 2: Centralized vs Decentralized Learning FL.

C. ENTROPY

Entropy is a central concept in information theory. It plays a fundamental role in understanding efficient communication and information transmission and providing a quantitative means of measuring uncertainty. Consider, for example, a simple system composed of coin flips. In such a system, both faces have an equal probability of occurrence. The entropy, denoted by $H(X)$, where X is a random variable that represents the result of each coin flip, can be calculated using the entropy formula. This calculation allows us to quantify the uncertainty of the information produced by the flips.

$$H(X) = -\sum_{i=1}^{n} p(x_i) \log_2 p(x_i) \tag{1}$$

Where:

i. X represents the set of all possible symbol values.
ii. $p(x_i)$ is the probability of occurrence of the i-th symbol value.
iii. The $\sum_{i=1}^{n}$ all possible values.
iv. $\log_2 p(x_i)$ is the logarithm with the base b of probability $p(x_i)$, making the entropy unit in bits.

In this case, there are heads and tails, where the entropy $H(X)$ represents a value of 1; thus, each coin provides one bit of information. However, with such a result, it is

impossible to precisely determine the outcome because each flip has the same degree of uncertainty.

However, if entropy is reduced, each new flip will be less surprising as outcomes become increasingly predictable. Conversely, uncertainty and unpredictability rise at maximum entropy, making system outcomes completely random.

Figure 3 presents the concepts discussed concerning entropy, highlighting its applicability as a metric for quantifying the uncertainty or the degree of unpredictability of an information source. It measures disorders and impurities in datasets, helping select relevant data features.

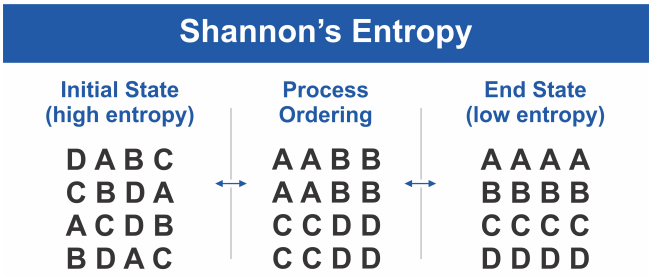


FIGURE 3: The figure illustrates entropy reduction through symbolic states: high, intermediate, and low entropy, representing decreasing uncertainty.

Additionally, entropy serves as a tool for analyzing how noisy or incomplete the data are, which, in turn, hinders the identification of patterns and negatively affects the convergence of ML models.

Furthermore, entropy contributes significantly to in ML, particularly in decision tree classification algorithms, where it measures the degree of disorder and impurity in the dataset to minimize the uncertainty regarding the classes in each split. The information gain criterion, derived from entropy, is employed to select the best data features, aiming for more accurate classification [35]. About data quality, entropy is also used in ML; noisy or incomplete data tend to increase a system's entropy, making it more difficult to identify clear patterns and reducing the efficiency of AI models [36].

III. RELATED WORK

Recent advances in FL have caused significant changes in communication efficiency, precision, and convergence optimization, particularly when dealing with the challenges imposed by non-iid and heterogeneous environments. These issues are summarized in Table 1, which presents the FL approaches related to this study and their main characteristics.

For instance, "FedAVO," a method inspired by natural optimization strategies to improve communication efficiency in FL, reflects an interest in solutions inspired by nature [37]. Likewise, "Fedco," which utilizes grouping optimization to increase communication efficiency, has been suggested to manage and effectively reduce data communication overload [38]. Orlandi et al. [39] presented the FedAvg-BE algorithm, which reduces the runtime in non-iid data in FL by up to 22% for MNIST and 26% for CIFAR-10 using edge entropy evaluation.

These advancements focus on improving communication efficiency, accuracy, and convergence in FL environments using non-iid data. Despite these advances, computational cost, communication, optimization, and data handling challenges in FL architectures remain crucial.

Some algorithms improve learning efficiency and precision, such as Yu et al. [20], which automatically adjusts weights for the best performance. Preconditioned FL was also introduced in this context, proposing a method to precondition learning environments or data to improve FL performance [40].

A subsequent study introduces "FedWNS," which utilizes the selection of nodes based on data distribution through learning by reinforcement, highlighting a node selection strategy to obtain better results [21]. In another approach, Wolfrath et al. considered heterogeneity and focused on selecting grouped clients, accelerating the FL process to address the challenges of client heterogeneity during the learning process.

In the Li et al. [41] approach, clients send their models to the server and share the distribution of their training data, enriching the server with additional information and global optimization. However, this last approach violates the sensitive data strategy of the FL.

Other approaches have introduced the FL architecture based on blockchain technology. It leverages a data and model provenance ledger built on intelligent contracts and a fair and weighted data sampling algorithm [42]. Similarly, incentive mechanisms have developed to encourage node participation thoughtfully, increase participation, and ensure collaboration more fairly [43].

Yang et al. [44] proposed an aggregating strategy that improves the model's convergence speed in non-iid environments by accounting for server-side characteristics with high variation. Similarly, Dolaat et al. [45], and Xu et al. [46] introduced strategies to enhance precision and personalize global FL models to address non-iid challenges. These approaches emphasize incentive mechanisms, aggregation techniques, and balancing strategies in FL environments.

Ma et al. [12] explored the efficacy problem of AI model training in distributed scenarios, especially the solution to non-iid data in FL, and the relevance of efficient data selection methods. Parallel to this, other authors approach methodologies to improve convergence and broach data heterogeneity through local model training adjustments and "hyper-knowledge" sharing [19], [47].

Researchers have introduced mechanisms to adjust gradients, optimize learning in environments with data heterogeneity, and optimize global structures, approaching critical challenges in the search for an efficient FL [23], [48]. These studies demonstrate the impact of data heterogeneity and how adjustments in local models, particularly at the edge,

remain one of the literature's most promising and central focuses.

The FL convergence improvement through regularization was recently proposed by Qiao et al. [49] to address the data heterogeneity between clients directly. Additionally, the FedGroup Framework incorporates a client clustering strategy using the K-means++ algorithm and optimization techniques, including meta-learning, adaptive optimization, and gradient aggregation strategies [50]. Another study by Ilic et al. [51] simultaneously examined several clients updating a global model.

The experiments involved aggregated updates using a method known as federated averaging. Incremental updates, in which the global model undergoes sequential updates, are also considered. This study also includes cyclic updates, where minor updates occur per epoch, and semi-simultaneous updates, combining simultaneous and incremental strategies. Moreover, these studies highlight the importance of advanced strategies at nodes (clients), where these techniques significantly affect global models' learning processes and updates.

Noise in the data under non-iid scenarios poses an additional challenge for global FL models. In this context, [58] proposed "FEDCNIA," which is an approach aimed at mitigating the impact of noise on FL clients. Furthermore, the discrimination in data distribution in FL [59] is explored using techniques to handle the non-iid nature of data in the healthcare and finance sectors. Regarding client data heterogeneity, [24] proposed an adaptive mechanism inspired by the Shapley value to promote greater client fairness. Client consistency is a critical factor affecting aggregation models [22].

Thus, the FEDLAW algorithm demonstrates how data heterogeneity, the number of local epochs, and variability among clients influence the global model [62]. However, noise and data heterogeneity remain critical challenges in global aggregation techniques. The proposed model aims to mitigate quality loss and address data discrimination to ensure fairness among clients and quality during training.

Finally, several studies in the medical field have explored the application of FL, emphasizing different aspects and challenges. For example, Antunes et al. [63] highlighted open questions regarding adoption and data aggregation mechanisms in Electronic Health Records (EHR). In medical imaging, the use of FL focuses on privacy concerns regarding brain tumors [45]. Furthermore, technical challenges of FL, such as non-iid data, were discussed in [50] using heart rate data.

Additionally, [11] addresses issues related to complications in data transfer in the healthcare field and other medical data and applications. Furthermore, several studies have highlighted the applications of FL, emphasizing its importance in data privacy, sensitivity, and global aggregation models. These issues in the medical field address the technical challenges related to the heterogeneity of health data and the area of medical imaging.

A. PROBLEM OVERVIEW

In this section, we explore the main topics of the problems addressed in this work and the state-of-the-art literature, focusing on challenges related to data quality in FL, particularly in non-iid scenarios.

The exponential growth of IoT devices and the massive volume of generated data pose significant challenges for practical analysis and the extraction of meaningful insights while addressing critical issues such as privacy protection, statistical heterogeneity, optimization and performance, and communication efficiency, all while preserving privacy and accuracy [64], [65].

These are especially critical scenarios where devices face battery limitations, communication costs, latency, and synchronization issues.

In the context of FL, data exhibits complex characteristics, such as non-iid properties, necessitating efficient algorithms and architectures capable of securely processing large data volumes, respecting privacy, and enhancing performance.

Thus, the central problem addressed in this work is data quality, mitigating unnecessary processing by eliminating redundant information that does not add value to the neural network or contribute significantly to model accuracy.

The goal is to address device-related issues and reduce energy costs. This approach allows AI systems to train more rapidly while consuming fewer computational resources.

This work tackles these challenges by proposing an algorithm to improve data quality in the FL context, focusing on identifying information that efficiently contributes to the neural network and reducing the computational cost of training.

B. DISCUSSION OF PROBLEMS

The FL model inherently exhibits high latency, causing delays in convergence, reduced update frequency, synchronization issues, impact on accuracy, and increased energy consumption. Furthermore, one of the current limitations of the federated environment is the selection of nodes that frequently leave the system and are replaced by new nodes that incrementally introduce information.

In this context, the input ignores the quality of the selected data, focusing instead on the processing capacity and availability of the nodes (clients). The data that contribute minimal incremental value to the model result in consequences at the edge processing level. As a result, such data is underutilized and fails to produce a significant update in the local model weights, leading to unnecessary computation at the edge, excessive communication requirements, and increased energy consumption.

Since none of these new node sets are specifically validated, they allow for input data of varying quality, which can potentially increase latency and make it difficult to achieve high accuracy and low loss.

By contrast, the proposed model (e.g., Entropy-Based Selection (EnBaSe)) prioritizes both the nodes' quality and

TABLE 1: Overview of Federated Learning Studies by System Characteristics and Optimization Techniques.

Author	Date	Properties				Strategies							
		Deep Learning	Machine Learning	Edge Computing	IoT Devices	Algorithm Optimization	Automatic Adjustment	Data/Client Selection	Data Distribution	Image Data Analysis	Data Quality Analysis	Aggregation Method Compatible	Embedded Compatible
Li, Beibei et al. [52]	2020	x				x				x			
Kang, Jiawen et al. [16]	2020	x											
Du, Zhaoyang et al. [53]	2020	x		x	x				x				
Itahara, Sohei, et al. [18]	2021	x	x		x	x		x	x				
Criado, Marcos F. et al. [10]	2022	x	x						x				
Gafni, Tomer et al. [54]	2022	x		x	x	x			x				
Al-Saedi, Ahmed A et al. [38]	2022	x		x	x	x			x	x			
Yu, Xi et al. [20]	2022					x	x		x	x			
Ullah, Shan et al. [47]	2022	x				x			x	x			
Xu, Jian et al. [46]	2022	x				x			x	x			
Wolfrath, Joel et al. [55]	2022	x		x	x	x		x	x	x			
Li, Yang, et al. [56]	2022	x								x	x		
Zhang, Yu, et al. [35]	2023	x	x			x				x			
Condori Bustincio, et al. [25]	2023	x	x		x			x	x				
Orlandi, Fernanda C. et al. [39]	2023	x		x	x	x	x	x	x	x			
Lo, Sin Kit et al. [42]	2023	x		x	x	x		x	x	x			
Hossain, Md Zarif et al. [37]	2023			x	x	x			x	x			
Tao, Zeyi et al. [40]	2023	x		x	x	x			x	x			
Lee, Hyeongok et al. [57]	2023					x			x	x			
Tu, Chengwu et al. [21]	2023	x		x	x	x		x	x	x			
Li, Boyuan et al. [41]	2023	x		x	x	x			x	x			
Yang, Wei-Jong et al. [44]	2023								x	x			
Chen, Huancheng et al. [19]	2023	x		x	x	x			x				
Zheng, Shu et al. [23]	2023					x			x	x			
Huang, Chenxi et al. [48]	2023	x				x			x	x			
Dolaat, Khalid Mahmoud Mohammad et al. [45]	2023		x	x	x	x			x	x			
Qiao, Yu et al. [49]	2023			x	x	x			x	x			
Sabah, Fahad et al. [50]	2023	x	x	x				x	x				
Li, Zexi et al. [22]	2023	x	x			x		x	x	x			
Wu, Chenrui et al. [58]	2023				x		x		x				
Sun, Qiheng et al. [24]	2023					x		x	x	x			
Iyer, Venkataraman Natarajan et al. [59]	2024		x						x	x			
Milan Ilić et al. [51]	2024	x	x	x		x							
Yan, Litao, et al. [60]	2024		x	x	x	x					x		
Hamidi, Shayan Mohajer, et al. [61]	2024	x	x		x	x	x	x	x				
Our Model	2024	x	x	x	x		x	x	x	x	x	x	x

processing capacity. Thus, the EnBaSe entropy algorithm excludes information that does not significantly contribute to the model or might have a limited contribution, selecting data with the most relevant information.

Proper validation of the data's quality can increase its homogeneity. For example, suppose the average entropy of a data set is reduced from 4.8 to 4.6215, representing a decrease of 0.1785. This reduction means less uncertainty and unpredictability in the system, messages, or processed data.

When reduced, entropy, which measures global uncertainty, implies greater predictability and uniformity. Therefore, this results in images and data with a high standard of consistency, simplifying the models and eliminating inconsistent samples from the system. Removing these sample inconsistencies speeds up the convergence of the neural network, allowing the model to be trained more quickly and efficiently. Consequently, this reduces energy consumption and processing, optimizing the system's performance.

For clarity, the MNIST dataset, comprised of 60,000 images in the iid scenario, has a total of 4.8 bits of entropy. Reducing entropy by removing 30,000 images equates to an increase in accuracy from 98.95% to 99.27%, as shown in Table 9, increasing the system's predictability and reducing the computing cost of processing 30,000 images that added no value to the neural network.

For the non-iid scenario, following the same pattern of data removal, this results in a loss of accuracy from 81.73%

to 81.26%, as shown in Table 12. Consequently, the removed images add zero or negligible value (0.47%) to the neural network's accuracy, wasting processing time and energy.

IV. PROPOSED MODEL

This study tests the hypothesis that it is possible to use information theory to quantify and analyze the quality of information in a dataset, as well as the uncertainty or surprise associated with its data distribution. Additionally, improving the quality of the input data and minimizing data noise were evaluated to reduce the computation time and energy costs.

In line with this perspective, the current state-of-the-art research explores the application of Entropy to quantify the degree of uncertainty and assess the redundancy present in information. These studies aim to comprehensively analyze information systems to measure the informational gain achieved through data processing.

In this context, Entropy is a tool for identifying data portions that effectively contribute new information to a system. Analyzing redundancy enables a more detailed evaluation of distortions, data quality, and information reliability.

This approach reduces unpredictability, offers a robust framework for quantifying information with low bias, and becomes a metric of information gain [56], [66]. Based on these principles, Entropy is approached as a control mechanism in deep neural networks to address the heterogeneity of data and clients [39], [61].

Furthermore, it is argued that Entropy is a suitable metric to assess the degree of disorder in a system (e.g., dataset) [67]. As a measure of disorder, capturing this fundamental characteristic of the system is considered a suitable approach. Entropy helps identify and quantify heterogeneity between clients' data, allowing for adaptive adjustments to achieve better convergence of the global model.

In addition, Entropy is also used to identify subsets of relevant and representative data. By serving as a metric to assess the relevance or diversity of data, it ensures that clients have meaningful and relevant information [25]. Additionally, some studies argue that Entropy reduces communication overload, as only relevant clients send updates, decreasing the required communication [18], [39], [60].

A. HYPOTHESIS

The hypothesis is that when Entropy applies to the field of Computer Vision (CV), each set of pixels begins to represent the color or intensity values of a specific color or intensity. Therefore, the probability of each color or intensity value and each unique pixel occurrence can be calculated based on the number of times each specific color or intensity appears in the image.

Thus, high Entropy represents a great diversity of pixels, indicating a high complexity in texture, significant variation, and little predictability of the information. On the other hand, low Entropy indicates greater image homogeneity, that is, better uniformity in identifying regions with little or no relevant information, facilitating the segmentation of elements in a scene.

Finally, the assumption is that the available data adds little value to the model and introduces noise into the training process. The following analogy illustrates this idea: initially, a disorganized set of images presents high Entropy and great uncertainty. By organizing and separating these sets of images, they can be divided into segments with low Entropy, making them highly predictable.

In contrast, the other part exhibits high Entropy and continuous unpredictability. Consequently, data with high Entropy tends to be viewed as low-quality or noisy and is thus excluded from the training process to enhance model performance.

B. REFERENCE MODEL AND OPERATION SCHEME

Figure 5 presents a model of the algorithm with separate steps. Initially, Entropy is computed for each 2D image class, represented by matrices, forming key-value pairs, where the key is the number of images, and the value is the linked Entropy. Subsequently, these entropy values were organized sequentially in each class and divided based on a probabilistic distribution, utilizing the median as a criterion, resulting in selected classes.

Finally, the process sends the data to the neural network for training. It selects values below the median of each class and stores them as representatives. This step applies to all classes in the dataset.

Figure 4 illustrates the interaction between the model proposed in this section and scenarios with iid and non-iid data. Specifically, Figures 4 (**1**) and (**2**) demonstrate the centralized scenario where, in (**1**), data from centralized devices are used for training an AI model, and in (**2**), global parameters are sent by the devices for aggregation into a global model. Figure 4 (**3**) shows an intermediate layer that facilitates processing, aggregation, orchestration, device management, and security. Finally, Figure 4 (**4**) represents the dataset from different IoT devices that have been centralized or will be used in decentralized learning.

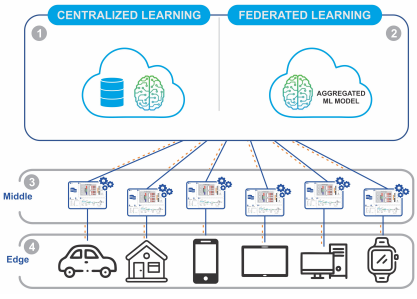


FIGURE 4: EnBaSe model applied to iid and non-iid scenarios.

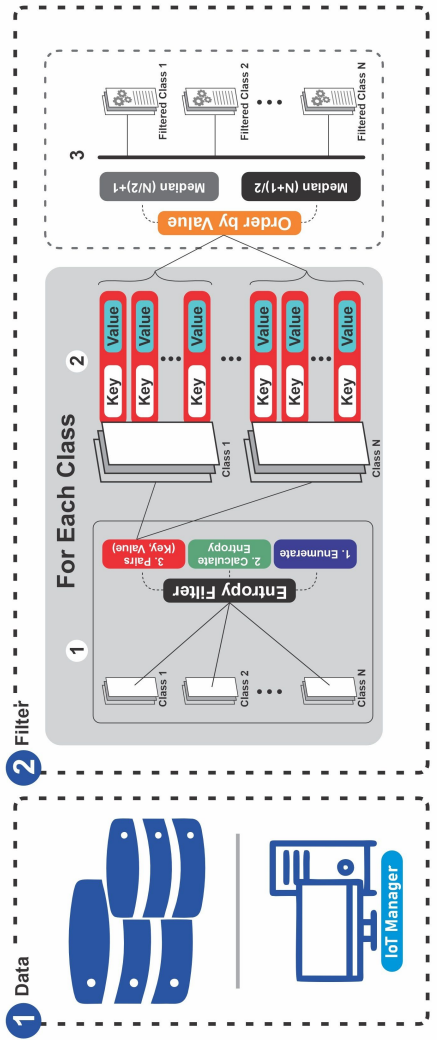


FIGURE 5: Algorithm Model: EnBaSe (Entropy-Based Selection).

C. ALGORITHM DETAILS

As understood in state-of-the-art studies in Information Theory, Entropy is a mathematical and statistical tool used to measure the degree of disorder and information gain in a system. Reducing the Entropy within an information system increases the predictability of outcomes. In this context, the proposed algorithm maps each system by selecting the most significant data within each subset.

Building on this concept, creating a data subset with reduced redundancy and noise is possible, which ensures a higher quality of the data subset. Consequently, neural networks can be trained more efficiently and at lower computational costs.

The Algorithm 1 is designed to be implemented in embedded systems, whether centralized servers or IoT devices. Specifically, among these devices is the most representative dataset from each data subset, using Entropy to identify the most informative and representative data.

As illustrated in Figure 5, the proposed model interacts with both the centralized and decentralized environments, as shown in Figure 4. Thus, the EnBaSe algorithm balances the selection of each dataset, selecting half of the dataset with the lowest Entropy for each class, aiming to create a homogeneous sample of each class and balance both the centralized and decentralized scenarios to reduce computational, energy, and time costs.

The algorithm is embedded and receives training sets and labels, represented by Initialization, which occurs by creating two sets for data storage: $\mathcal{X}$selected and $\mathcal{Y}$selected. The algorithm iterates over each class from 0 to K-1, calculating the Entropy and measuring the degree of disorder for each image in each class. Finally, it selects the data below the median and returns the selected data in $\mathcal{X}$selected and $\mathcal{Y}$selected, which serve as input for training. Algorithm 1 is as follows:

Algorithm 1 EnBaSe. Where K denotes the total number of classes.

Require: $\mathcal{X}_{\text{train}}$, $\mathcal{Y}_{\text{train}}$, K
Ensure: Selected classes based on Entropy
$\mathcal{X}_{\text{selected}} \leftarrow \emptyset$
$\mathcal{Y}_{\text{selected}} \leftarrow \emptyset$
for label $\leftarrow 0$ **to** $K-1$ **do**
  $\mathcal{C} \leftarrow$ Retrieve indices belonging to class label
  $\mathcal{M}_{\text{Entropy}} \leftarrow \emptyset$
  for each sample $\in \mathcal{C}$ **do**
    $\mathcal{M}_{\text{Entropy}} \leftarrow (\text{key}, \text{ComputeEntropy}(\text{image}))$
  Sort $\mathcal{M}_{\text{Entropy}}$ **by** ComputeEntropy(image)
  Calculate the median of $\mathcal{M}_{\text{Entropy}}$
  $\mathcal{I}_{\text{Qualified}} \leftarrow \emptyset$
  for each key $\in \mathcal{M}_{\text{Entropy}}$ **do**
    if key.entropy $\leq$ median **then**
      Append key.index to $\mathcal{I}_{\text{Qualified}}$
  for i **in** $\mathcal{I}_{\text{Qualified}}$ **do**
    Append $\mathcal{X}_{\text{train}}[i]$ to $\mathcal{X}_{\text{selected}}$
    Append $\mathcal{Y}_{\text{train}}[i]$ to $\mathcal{Y}_{\text{selected}}$
return $\mathcal{X}_{\text{selected}}$, $\mathcal{Y}_{\text{selected}}$
function COMPUTEENTROPY(image)
  $H \leftarrow -\sum_d p(\text{image}) \log_2(p(\text{image}))$
  return H

where

i. K: Represents the total number of classes in the dataset.
ii. $X_{\textbf{train}}$: Training Dataset.
iii. $Y_{\textbf{train}}$: Labels corresponding to the training set X_{train}.
iv. $X_{\textbf{selected}}$: Subset X_{train} selected by the algorithm based on entropy.
v. $Y_{\textbf{selected}}$: Labels corresponding to subset X_{selected}.
vi. **label**: A class label ($K-1$, where K is the total number of classes).
vii. **C**: Set of indices belonging to a given class $label$.

viii. **MEntropy**: An array that stores pairs (index, entropy value) for each image in a given class.
ix. **ComputeEntropy(image)**: A function that calculates the entropy of an image.
x. **Median**: Median entropy values in the set $MEntropy$.
xi. **IQualified**: Set of indices of samples with Entropy less than or equal to the median, X_{selected} and Y_{selected}.
xii. **H**: Entropy calculated for the image.
xiii. $p(image)$: Probability distribution associated with the image (used in entropy calculations).

Algorithm 1 operates through several stages, detailed as follows:

i. **Initialization**: Two empty sets ($\mathcal{X}_{\text{selected}}$ and $\mathcal{Y}_{\text{selected}}$), are created to store the selected data and their corresponding classes.
ii. **Iteration over classes**: The algorithm iterates through each class present in the training set ($\mathcal{X}_{\text{train}}$ and $\mathcal{Y}_{\text{train}}$), identifying the indices associated with each class.
iii. **Entropy computation**: For each element in the current class, the ComputeEntropy function calculates the Entropy of the sample based on the probability distribution of its attributes. The results appear as $(\text{index}, \text{ComputeEntropy})$ pairs in $\mathcal{M}_{\text{Entropy}}$, which is also referred to as the entropy map.
iv. **Sorting and median-based selection**: The $\mathcal{M}_{\text{Entropy}}$ pairs are sorted by Entropy, and the median entropy value is calculated. Only elements with entropy values less than or equal to the median are selected, ensuring the inclusion of the most representative and least redundant data.
v. **Updating the selected subsets**: The selected indices are used to copy the corresponding data from $\mathcal{X}_{\text{train}}$ and $\mathcal{Y}_{\text{train}}$ to the subsets $\mathcal{X}_{\text{selected}}$ and $\mathcal{Y}_{\text{selected}}$.
vi. **Final output**: At the end of the Iteration over all classes, the subsets $\mathcal{X}_{\text{selected}}$ and $\mathcal{Y}_{\text{selected}}$ contain the most informative and homogeneous data, optimized for model training.

In summary, the algorithm proposed in this study, Algorithm 1, leverages Entropy as a metric to identify relevant data subsets for each class, aiming to reduce redundancy and noise while enhancing the quality of the selected data. The choice of lower Entropy is grounded in Information Theory, which asserts that systems with lower Entropy exhibit greater predictability. Beyond its data selection strategy, the EnBaSe is designed to operate independently, ensuring broad applicability across different FL methods.

This design choice means that EnBaSe operates independently of an embedded design. The approach remains decoupled from any specific FL algorithm. Due to this, our algorithm is general-purpose and compatible with any global aggregation method.

This versatility is particularly beneficial in scenarios where computational resources are limited. Thus, the data subsets are structured homogeneously to optimize the training process. For instance, in medical classification systems with computational power constraints, Algorithm 1 prioritizes more informative and less redundant features, reducing the computational effort required by the neural network. This approach enables rapid and highly accurate responses in real-time IoT systems.

D. MATHEMATICAL FORMULATION OF ENBASE

The model described at the beginning of Subsection IV-B, based on Hypothesis IV-A, builds on the principles of Information Theory to extract data with higher informative quality II-C. In particular, it applies Shannon entropy to quantify the degree of disorder in each image, allowing the identification of low-entropy samples that best represent the neural network domain. The following formulations define the mathematical base of pseudocode in the Subsection IV-C.

Let the amostral space I represent a matrix of pixel values from an image, and $p(x_i)$ be its probability distribution of pixel values x_i in the image. Then, the Shannon entropy $H(I)$ is defined as:

$$H(I) = -\sum_{i=1}^{d} p(x_i) \log_2 p(x_i) \tag{2}$$

Entropy $H(I)$ quantifies the degree of uncertainty in new data: the more uncertain it is, the more associated information there is. Which is calculated using the logarithm $\log_2$ and measured in bits. Thus, $p(x_i)$ represents the frequency of occurrence of the pixel value x_i. A Low $H(I)$ value indicates a low degree of uncertainty in the image. Therefore, a low entropy value implies high predictability, which benefits neural network training when specializing in a specific subset of data. This approach allows the neural network to learn more efficiently with fewer input data.

Thus, given a class $c \in \{1, \ldots, K\}$ with a sample set $\mathcal{I}_c = \{I_{c_1}, I_{c_2}, \ldots, I_{c_n}\}$, the sorted set of image entropies is defined as:

$$\mathcal{H}_c = \left\{ H(I_{c_j}) \;\middle|\; j = 1, \ldots, n \right\} \tag{3}$$

The selected subset $\mathcal{S}_c$ from the class c is defined based on the median Entropy from $H(I_{c_j})$ set, and consists of the image collections whose Entropy is less than or equal to $\mathcal{H}_c$:

$$\mathcal{S}_c = \{ I_{c_j} \in \mathcal{I}_c \mid H(I_{c_j}) \leq \mathcal{H}_c \} \tag{4}$$

The data selected for training the neural network is represented by the union of selected subsets for all classes of $\mathcal{S}_c$.

$$\mathcal{S} = \bigcup_{c=1}^{K} \mathcal{S}_c \tag{5}$$

EnBaSe selects samples based on Entropy that is lower than or equal to the median value of $\mathcal{H}_c$ for each class c, thereby constructing a more informative subset $\mathcal{S}_c$. The

final training set $\mathcal{S}$ is obtained as $\mathcal{S} = \bigcup_{c=1}^{K} \mathcal{S}_c$, combining all subsets across the K classes. The higher predictability and lower noise of the data reduce the vanishing gradient problem and enable the model to learn more effectively with the data. As a result, the selected subset ($\mathcal{S}$) indirectly reduces computational costs and accelerates the convergence of the neural network models.

V. EXPERIMENTAL EVALUATION

This section provides a comprehensive outlook on the methodological approaches adopted during the execution of this experiment, encompassing the steps taken, enabling the replication of the experiment by other interested researchers, and facilitating a critical and detailed evaluation of the results obtained. Source codes are available in Github[1] for reproducibility.

A. DATASET DESCRIPTION

The MNIST, Fashion MNIST, CIFAR-10, and CIFAR-100 datasets are fundamental for the training and evaluating ML and CV. The scope of the experiment encompassed the same datasets as the iid and non-iid scenarios. The characteristics of the selected datasets are detailed as follows:

i. **MNIST:** images in a grayscale of handwritten digits (0-9), divided into training and test sets. The goal was to develop models for recognizing and classifying these digits.
ii. **Fashion MNIST:** Alternative to MNIST contains images of fashion articles in ten categories, such as shirts and pants, in grayscale. These images pose a challenge for accurate classification.
iii. **CIFAR-10:** Presents colored images in ten different classes, including cars and animals, with training and test sets utilized for benchmarking in image recognition.
iv. **CIFAR-100:** Similar to CIFAR-10, but with 100 classes for more significant granularity, ranging from people to natural elements, representing a challenge amplified by the increased number of classes.

The experiments evolved from MNIST, Fashion-MNIST, and CIFAR-10 to CIFAR-100 to validate the performance of FL methods in distributed and heterogeneous CV environments with EnBase Algorithm at the edge. Since modern edge devices handle high-dimensional visual data with significant class variation and complexity, for instance, the CIFAR-100 dataset realistically captures this environment in a standard dataset that enables the reproducibility of our experiments.

On the other hand, validating our results by researchers is difficult without standard datasets due to the impossibility of reproducibility from the heterogeneous input data source. **To mitigate this limitation**, the datasets were selected for the experiment **based on the state-of-the-art review in Table 14**, which presents the main continuous and discrete datasets used in the literature, as well as studies that exhibit a unimodal nature concerning the edge computing and CV challenges addressed in this experiment. This table outlines FL algorithms, aggregation approaches, application scenarios, metrics used, advantages, disadvantages, and the most recurrent datasets in the CV area applied to experiments in FL.

Recognizing that, compared to the real world, the results have inherent limitations, **we adopt strategies suggested in the literature to simulate more realistic scenarios**. In this context, the heterogeneity and distributional properties of the data are considered, as presented in Subsection V-C and according to [68], [69].

Therefore, we conclude that, **in this methodology**, the experiment strategy **is based on state-of-the-art recurrent datasets** (e.g., for FL in *edge* applications in the CV area), as well as on the **distributions** of the **datasets identified in the literature**.

B. IID EXPERIMENT CONFIGURATION

To address the different challenges related to image processing and pattern recognition in an iid setting, we chose four different databases to validate the method developed in this experiment. Table 2 lists the MNIST, Fashion-MNIST, CIFAR-10, and CIFAR-100 datasets used.

TABLE 2: Datasets Summary.

Dataset	Description	Format	Training/Test
MNIST	Hand-written Digits	28x28	60,000/10,000
Fashion-MNIST	Clothes	28x28	60,000/10,000
CIFAR-10	Assorted	32x32	50,000/10,000
CIFAR-100	100 Classes	32x32	50,000/10,000

We aim to utilize these datasets to address a range of benchmarks and challenges in CV, for MNIST and Fashion-MNIST, 20% of the data was designated for validation, whereas 10% was allocated for CIFAR-10 and CIFAR-100.

C. NON-IID EXPERIMENT CONFIGURATION

1) Data Distribution

For the creation of the non-iid scenario in federated learning, where the data is not independent and identically distributed, the model is based on feature distribution skew, label distribution skew, and quantity skew [68]; these types of skewed distributions imply characteristics of heterogeneous data [69].

The non-iid indices for the clients (nodes) are used as a quantitative metric of the degree of data distribution among the clients (nodes), focusing on the **feature distribution**, **label distribution skew**, and **quantity skew**. The key considerations are as follows:

i. **Distortion in Feature Distribution:** Feature distortion refers to the imbalance between different quantities of labels across various clients concerning a specific client, which significantly affects performance and training. This feature distortion results in each client

[1] https://github.com/ernesto-arq/Entropy-Artificial-Intelligence.git

(node) having different features that may correspond to the same label, which can contain different information. For instance, the same character can be written in various styles, such as stroke width or inclination variations, leading to heterogeneous representations of the same labels.

ii. **Distortion in Label Distribution:** Label distortion occurs when different clients (nodes) in distinct locations exhibit varied distributions owing to demographic differences. These variations result from demographic and contextual factors that affect each client's label occurrence frequency (node).

iii. **Quantity Distortion:** Quantity distortion refers to an imbalance in the number of specific labels within a client, which affects the amount of data available for a single client (node). This imbalance significantly fragments the training process and model performance, creating an under-representation or over-representation of specific labels, thereby affecting the balance of the model.

Initially, nodes are formed using random datasets based on available data, where sufficient or available data from specific classes are not guaranteed. This random distribution introduces heterogeneity across nodes because different nodes may receive varying amounts of data or data types.

The goal is to ensure that any variation between groups (nodes) results from a random rather than a systematic factor, thereby reflecting the inherent variability in IoT environments. Allowing different nodes to contribute unevenly captures a more realistic representation of the diverse data encountered in real-world IoT devices, enhancing the model's ability to generalize across heterogeneous conditions.

2) Aggregation Algorithm

Aggregation approaches are widely recognized in the literature and are well established within the field of federated learning and have been selected. This selection aimed to validate the hypothesis concerning entropy's capacity as a quantitative measure for evaluating data quality in a federated environment, thereby reducing latency, noise, computational cost, and energy consumption. This experiment employed the following global aggregation models:

- **FedAvg:** calculates the weighted average of each client update. In this calculation, the weights typically correspond to the volume of data each client possesses, thus adjusting for any imbalances in the dataset.
- **FedProx:** FedProx introduces a new approach to the local loss function by adding a proximal regularization term. This term penalizes significant adjustments in the local model weights that deviate significantly from the global model weights. Thus, FedProx seeks to minimize the impact of data and device heterogeneity, promoting more harmonious learning.

D. DEEP NEURAL NETWORKS

Experiments with iid: We analyzed the MNIST and Fashion-MNIST datasets without Transfer Learning (TL) or Data Augmentation (DA) using the Stochastic Gradient Descent (SGD) optimizer. MNIST had 32 batches, and Fashion-MNIST had 128 batches, both for ten epochs. In CIFAR-10, we applied DA and used the Adam optimizer with 128 batches for 50 epochs. For CIFAR-100, we combined DA and TL with SGD, trained for 50 epochs in batches of 128, and added a callback to improve training control.

To enhance the stability and performance of Deep Learning (DL) algorithms in centralized scenarios with homogeneous data, dataset pixel values were normalized to a 0–1 scale by dividing them by 255.

The model configurations are listed in Table 3. For a fair comparison of the performance of the proposed algorithm, the same architectures were used with the complete dataset, thus allowing an assessment of its performance and limitations when trained with all the data.

TABLE 3: Model Configurations iid.

MNIST	Fashion	CIFAR-10	CIFAR-100
CONV-1	CONV-1	CONV-1	ResNet50
MP	MP	BN	LAYER-1
CONV-2	DP-1	CONV-2	LAYER-2
DP-1	FC-128	BN	LAYER-3
BN	DP-2	MP	LAYER-4
MP	FC-10	DP-1	GAP
CONV-3	SOFTMAX	CONV-3	FL-512
BN		BN	BN
FC-128		CONV-4	DP-2
FC-10		BN	FL-100
SOFTMAX		MP	SOFTMAX
		DP-2	
		CONV-5	
		BN	
		CONV-6	
		BN	
		DP-3	
		FL-512	
		BN	
		DP-4	
		FL-10	
		SOFTMAX	

Note: BN: batch normalization; MP: max pooling; DP: dropout; FC: fully-connected layer; GAP: global average pooling. Layer 1: three blocks (Conv, BN, ReLU, MP); Layer 2: four blocks (Conv, BN, ReLU, MP); Layer 3: six blocks (Conv, BN, ReLU, MP); Layer 4: three blocks (Conv, BN, ReLU, MP).

Experiments with non-iid: For the MNIST and Fashion-MNIST datasets, Convolutional Neural Network (CNN) were used without applying TL or DA, optimized by SGD. For CIFAR-10 and CIFAR-100, the adapted ResNet-50 model was employed, incorporating DA and optimized with SGD. All models were trained for 50 epochs with batches of 128 examples.

Different normalization parameters for the FL scenario with heterogeneous data were applied to normalize MNIST, Fashion-MNIST, CIFAR-10, and CIFAR-100 datasets for MNIST, an average of 0.1307 and a standard deviation of 0.3081. In the Fashion-MNIST dataset, the average

and standard deviation of the parameters are 0.2860 and 0.3530, respectively. For CIFAR-10, the average values were 0.4914, 0.4822, and 0.4465, with standard deviations of 0.2023, 0.1994, and 0.2010, for the **Red, Blue, Green (RGB)** channels, respectively. Finally, for CIFAR-100, the averages were 0.5071, 0.4867, and 0.4408, and the standard deviations were 0.2675, 0.2565, and 0.2761, respectively, for the RGB channels. The model configurations are presented in Table 4 following previously explained criteria.

TABLE 4: Model Configurations non-iid.

MNIST	Fashion	CIFAR-10	CIFAR-100
CONV-1	CONV-1	ResNet50	ResNet50
MP	MP	BN	BN
CONV-2	CONV-2	LAYER-1	LAYER-1
MP	MP	LAYER-2	LAYER-2
FL-500	FL-500	LAYER-3	LAYER-3
FL-10	FL-10	LAYER-4	LAYER-4
SOFTMAX	SOFTMAX	GAP	GAP
		FL-512	FL-512
		BN	BN
		DP-1	DP-1
		FL-10	FL-100
		SOFTMAX	SOFTMAX

Note: BN: batch normalization; MP: max pooling; DP: dropout; FC: fully-connected layer; GAP: global average pooling. Layer 1: three blocks (Conv, BN, ReLU, MP); Layer 2: four blocks (Conv, BN, ReLU, MP); Layer 3: six blocks (Conv, BN, ReLU, MP); Layer 4: three blocks (Conv, BN, ReLU, MP).

E. EVALUATION METHODS AND METRICS

This section evaluates the methods used to identify unbiased and high-quality neural networks. The evaluation process is structured as follows.

- **Dataset Utilization:** Initially, the complete dataset is utilized to analyze performance, providing a comprehensive overview of model capabilities (in the experiment named All Data).
- **Random Selection Method:** A random selection method was applied to ensure diversity and impartiality during training. This technique is well-established in the literature. It is considered an essential step towards creating an adaptable and unbiased network to prevent overfitting and ensure accurate responses to new data challenges (in the experiment named Random).
- **Model Comparison:** The effectiveness of the trained models was assessed by comparing their performance across three distinct scenarios: using all available data, applying a 50% random selection of data, and employing our proposed algorithm (EnBaSe) that selects half of the dataset, which uses 50% of the data.

Subsequently, the evaluation employed the following metrics for a comprehensive comparison:

- **Accuracy:** Accuracy measures the proportion of correct predictions from the total samples, calculated by dividing the hits by the total number of samples.
- **Recall:** Recall evaluates the proportion of correctly identified true positives, calculated by dividing accurate positive numbers by the total positive numbers plus false negatives.
- **F1-Score:** The F1-Score is the precision and recall harmonic value, indicating equilibrium between them. Higher values indicate a better model performance. In this specific case, it will be used as an additional metric in the context of FL owing to the high heterogeneity of the data.
- **Loss:** The loss function measures the error between predictions and typical results, with specific functions for each problem (e.g., cross-entropy for classification). The goal is to minimize such losses to improve the model.
- **Learning Curve:** Represents model performance over time, comparing training and validation to detect over- and under-fitting and verifying model convergence.

VI. EVALUATIONS AND RESULTS

A. DATA ANALYSIS

To validate our hypothesis regarding the role of entropy in information quality selection, we analyzed its behavior and examined how it was affected by normalization and DA operations. Subsequently, we assessed data normalization's impact (scaling values 255 to match pixel sizes) on sample selection by comparing entropy before and after normalization.

The results demonstrated that entropy remained stable, with an average similarity of 14 decimal places. These results indicate that, despite changes in absolute values, the probabilistic importance of the data—and, therefore, its entropy—remains unchanged.

As expected, based on Information Theory, it is important to note that entropy, by definition, is a measure of uncertainty and disorder, and its invariance is an expected outcome, as entropy focuses on relative probabilities rather than absolute values. Table 5 presents the datasets before and after normalization.

TABLE 5: Entropy Comparison (MNIST).

Image Indexes	Before Normalization	After Normalization
27582	5.150483033018236	5.150483033018237
5760	5.242222088792437	5.242222088792438
29284	5.247872784912092	5.247872784912093
4484	5.291615062825755	5.291615062825755
18188	5.378906867665029	5.378906867665029
Image Indexes	**Before DA**	**After DA**
57362	1.4863968283704654	1.4863968283704652
37920	1.4966680960341756	1.4966680960341756
21618	1.5087870622820911	1.5087870622820911
29180	2.0933393541022376	2.0933393541022376
3637	2.0933393541022380	2.0933393541022376

Following a similar approach, the experiments applied

linear geometric transformations through DA techniques, during which the entropy value remained constant. These results confirm that entropy is determined by the intensity levels of the pixels, irrespective of their spatial position. The geometric transformations applied in DA, such as rotations and translations, preserve the data structure and modify only the position of the pixels while keeping their intensity values unchanged.

Such findings conclude that entropy is unaffected by geometric transformations, ensuring its relevance in areas where preserving specific informational properties is essential.

B. ENTROPY DISTRIBUTION AND SAMPLE SELECTION IN DATASETS

This study also explored the entropy behavior of the histograms. Thus, the entropy value of each image was calculated, and a histogram was created to analyze the entropy behavior in the dataset, as shown in Figure 6. The data were analyzed according to iid and non-iid scenarios. The study also investigated the uneven and imbalanced data distribution in assorted clients using statistical tests, such as Shapiro-Wilk, Kolmogorov-Smirnov, D'Agostino, and Pearson tests, to analyze entropy characteristics.

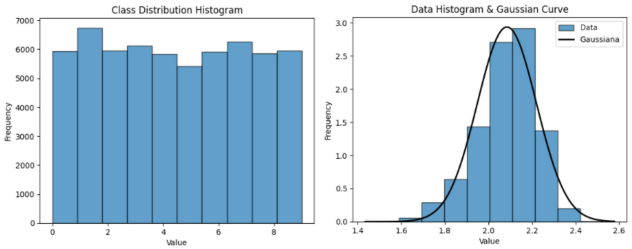


FIGURE 6: Distribution of Classes and Data Distribution after Entropy Calculation for MNIST.

During the development and experimentation of the EnBaSe algorithm, we found that the entropy values of each image, once calculated, sorted, and plotted in a histogram, exhibit a distribution that closely resembles a Gaussian distribution. This pattern holds for the MNIST, Fashion-MNIST, CIFAR-10, and CIFAR-100 datasets where entropy was applied.

Table 6 presents the results obtained from an average of five experiments for the distribution analysis, where we simulated 400 nodes receiving data in a distributed manner. In addition, we developed a consensus algorithm, as detailed in the table. This algorithm relies on a voting mechanism that deems the distribution normal when at least two-thirds of the remaining statistical tests confirm such characteristics.

i. **MNIST (iid & non-iid):** The class distribution in MNIST is homogeneous and balanced in the iid environment. In contrast, in non-iid, after entropy selection, it is observed that the data in a large part of the clients show a near-normal distribution. Entropy plays a significant role in bringing the distribution closer to a Gaussian curve.
ii. **Fashion-MNIST (iid and non-iid):** In Fashion-MNIST, after selection, the Gaussian distribution presents itself less uniformly. In a considerable amount of clients, the data follow a normal distribution.
iii. **CIFAR-10 (iid & non-iid):** In the CIFAR-10 set, a small proportion of the clients displayed normally distributed data after selection, indicating a less robust statistical distribution.
iv. **CIFAR-100 (iid & non-iid):** For CIFAR-100, a minimum number of clients displayed typically distributed data. This set was the most challenging in terms of approaching a Gaussian distribution.

TABLE 6: Statistical test results for different dataset scenarios.

Dataset	Consensus	Shapiro-Wilk	Kolmogorov-Smirnov	D'Agostino & Pearson
MNIST	344	344	399	342
Fashion-MNIST	181	181	391	223
CIFAR-10	37	37	318	71
CIFAR-100	7	7	203	17

Image datasets in state-of-the-art literature are composed of levels of challenge complexity, presenting more straightforward and more difficult labels for training AI models. From information theory, we know that data can possess high or low entropy, and consequently, we observe a distribution pattern that approaches a Gaussian distribution.

Therefore, this reinforces the hypothesis that, for images, there is a dataset of complex pictures with high variability and low predictability. Similarly, some data have low variability and high predictability for low entropy. As such, the EnBaSe model was constructed following an empirical and experimental approach based on information theory and state-of-the-art studies on entropy, utilizing low entropy and data close to the central position of the distribution.

Finally, the same observation supports the selection of a more robust metric, such as the median. Although these results approximate a normal distribution, one cannot assume that all datasets strictly follow the same pattern. The asymmetric distributions identified in this experiment indicate that the mean may shift towards the tails, potentially causing underestimation or overestimation.

C. ENTROPY COMPUTATIONAL COST

We conducted a total of 120 iid experiments (10 simulations for All Data, 10 for EnBaSe, and 10 for Random). In the non-iid scenario, we performed 240 experiments: 10 simulations each for the complete set, EnBaSe, and Random in FedAvg, and another ten simulations each for the complete set, EnBaSe, and Random in FedProx.

The entropy calculation times for the execution of the image-selection (EnBaSe) process are listed in Table 7, along with the average times for iid and non-iid environments, indicating the time required to compute the entropy

of the entire dataset and select the samples.

TABLE 7: Average Time for Entropy Computation.

Dataset	Distribution	Average (s)
MNIST	iid	≈1.76
Fashion MNIST	iid	≈1.83
CIFAR-10	iid	≈3.64
CIFAR-100	iid	≈3.67
MNIST	non-iid	≈2.77
Fashion MNIST	non-iid	≈3.21
CIFAR-10	non-iid	≈5.57
CIFAR-100	non-iid	≈5.49

Note: Average client (node) time in non-iid configuration; Centered mean time of dataset in iid configuration.

In addition, we highlight the use of Google Colaboratory (GC) for simulations in a Cloud Computing (CC) environment. We employed the T4 architecture, featuring high-speed memory, 12 Gigabyte (GB) of Random Access Memory (RAM), 15 GB Graphics Processing Units (GPU), and a 201.2 GB disk. We trained the iid scenario in a centralized environment using Scikit-learn, reflecting the sequential nature of the operations without advanced parallelization. In contrast, we used TensorFlow for the non-iid scenario.

D. EVALUATION IN THE IID SCENARIO

This section discusses the experimental results obtained under the iid scenario. Thus, we can evaluate the hypothesis that entropy influences the training process as a measure of data quality. The results include analyses of the MNIST, Fashion-MNIST, CIFAR-10, and CIFAR-100 datasets across various configurations, focusing on the data distribution, accuracy, recall, and error rate.

The table 8 presents the datasets used in the experiment for the iid scenario within a centralized architecture. These datasets, widely recognized in the literature, play a pivotal role in advancing ML as they enable the evaluation of deep neural network performance under diverse challenges and varying levels of complexity.

The MNIST dataset provides an essential starting point for model development because of its simplicity and widespread adoption in introductory studies. Conversely, the Fashion-MNIST dataset increases the complexity compared to MNIST, encompassing more diverse and challenging visual data. Meanwhile, the CIFAR-10 and CIFAR-100 datasets represent significant challenges as they involve image classification in more varied and complex scenarios, necessitating models with greater generalization capabilities.

TABLE 8: Datasets in centralized architecture.

Dataset	Description	Classes	Distribution	Architectures
MNIST	Handwritten digits	10	iid	Centralized
Fashion-MNIST	Clothing items	10	iid	Centralized
CIFAR-10	Various objects	10	iid	Centralized
CIFAR-100	100 categories	100	iid	Centralized

We distributed 120 experiments equally among the All Data, Random, and Entropy categories, as shown in Figure 7. In the MNIST dataset experiments, we focused on using entropy to guide data selection for this research. We conducted 30 experiments: 10 using the entire dataset, 10 with random selection, and 10 employing the proposed EnBaSe algorithm.

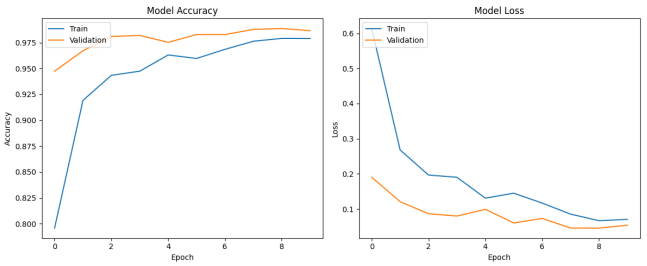


FIGURE 7: EnBaSe MNIST Training.

Figure 7 shows that the training and validation curves of EnBaSe are close, and when the validation curve surpasses the training curve, it indicates better generalization. A higher validation curve suggests that the model may benefit from regularization, thereby improving its generalization to new data.

The increase in the loss curve and reduction in the learning curve suggest that the model is starting to memorize the training data, indicating the need to adjust the learning rate to prevent deeper layers from learning less helpful patterns.

In the study involving the Fashion-MNIST dataset shown in Figure 8, we applied EnBaSe and presented a learning curve. We found a consistently higher validation curve than the training curve for the Fashion-MNIST dataset. Therefore, this suggests that the model effectively generalizes to unseen data, highlighting potential areas for improvement during the training process.

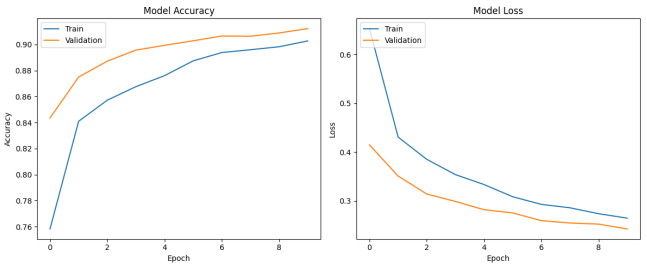


FIGURE 8: EnBaSe Fashion-MNIST Training.

Finally, we observed a positive and steady evolution in the learning curve, accompanied by a continuous decrease in loss for both the training and testing data. Consequently, this indicates that with each epoch, the model improves its ability to minimize the loss function and becomes progressively more accurate.

In Figure 9, the overlay of the training and validation curves, with only slight fluctuations where they intersect, generally indicates positive performance. As a result, the model consistently performed well on both the training and validation datasets. Their proximity suggests that the model is efficiently generalized, effectively transferring the knowledge acquired during training to the validation data.

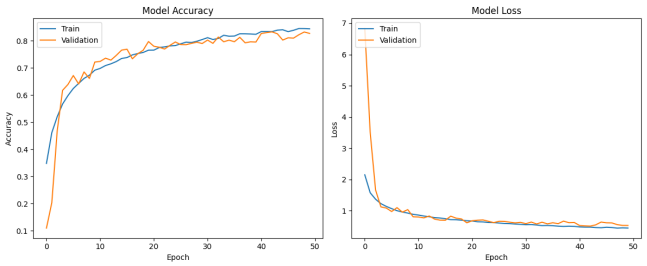


FIGURE 9: EnBaSe CIFAR-10 Training.

This scenario indicates a balance between bias and variance. A low bias reveals that the model can understand the complexity inherent in the data. Simultaneously, a low variance suggests that the model does not overfit the training data, allowing for good performance on the new data.

The learning curve, with a constant and high learning rate, and the loss curve, with reduced values, indicate that the model has reached or is close to its maximum potential within the constraints of its architecture and training settings.

As illustrated in Figure 10, the validation curve initially starts above the training curve, likely because of the composition of the training and validation sets, which helps the model learn more effectively. As the training progressed, the curves converged, reflecting the ability of the model to optimize its learning without sacrificing generalization.

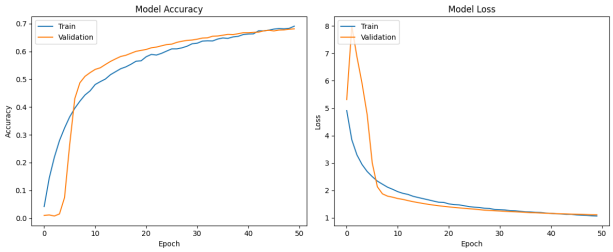


FIGURE 10: EnBaSe CIFAR-100 Training.

The decline in the training and validation losses over time confirms that the model learns effectively and avoids overfitting. For CIFAR-100, a well-known benchmark in image recognition, the EnBaSe algorithm exhibited a strong performance in handling the complexity of the dataset while preserving the model's generalization capability. For this reason, this highlights the effectiveness of EnBaSe in achieving efficient learning without introducing a bias that compromises the model's performance.

Table 9 presents the EnBaSe algorithm, which seeks to improve data quality by emphasizing consistency and precision, retaining only quality data, and reducing computational cost. We compared this algorithm with a random selection method, which tends to minimize selection bias and make the model more robust and less dependent on specific features.

In doing so, observing the general scenario of complete datasets and entropy behavior about a robust technique is possible. For a fair comparison, half of the set was used as a random dataset.

Furthermore, it presents the average results of the experiments with All Data, EnBaSe, and Random, which resulted in a minimal reduction in the accuracy for MNIST. For Fashion-MNIST, we observed the same pattern of accuracy. In CIFAR-10, the same accuracy pattern was observed.

Finally, for CIFAR-100, a benchmark in the CV field with 100 classes, a minimal and acceptable loss in processing cost savings was observed. In addition, a reduction in the overall computational cost for MNIST, Fashion-MNIST, and CIFAR-10 was evident without a significant compromise in accuracy.

As shown in Table 9, the EnBaSe algorithm demonstrated robust and consistent results when selecting half of the datasets, preserving the accuracy with minimal possible loss of quality. Comparing the results obtained with the accuracy reported in other state-of-the-art studies, it is evident that the algorithm is robust, highly scalable, lightweight, and can be easily integrated into embedded systems.

Furthermore, the algorithm exhibits high adaptability and can address various challenges in the field of CV using different architectures. Its efficiency is also evidenced by the reduction in computational costs and the acceleration of the convergence time of the AI model, enabling faster responses to events and dynamic environments, where time and computational cost are critical factors. For example, onboard health systems in the AI model can take significant time to acclimate to a patient's patterns.

Thus, we conclude that in the scenario with centralized data and iid, the EnBaSe algorithm is a computationally efficient solution that optimizes the selection of samples from the dataset with application in the field of CV. Its integration into embedded systems allows it to be applied dynamically as an AI tool in different centralized learning systems.

Table 10 presents a comparison with the available data from other studies that utilized similar neural network architectures, including the training and validation sets provided by the authors in recent studies, representing the state-of-the-art.

TABLE 9: Average results of the iid experiment.

Dataset	Type	Training			Validation			Test	Time (s)
		Accuracy (%)	Recall (%)	Loss	Accuracy (%)	Recall (%)	Loss	Accuracy (%)	
Mnist	All Data	≈ 99.53	≈ 99.52	≈ 0.038	≈ 98.95	≈ 98.94	≈ 0.049	≈ 99.19	≈ 73
Mnist	**EnBaSe**	≈ 99.28	≈ 99.27	≈ 0.070	≈ 98.64	≈ 98.61	≈ 0.050	≈ 97.61	≈ 32
Mnist	Random	≈ 99.49	≈ 99.49	≈ 0.055	≈ 99.49	≈ 98.77	≈ 0.058	≈ 98.93	≈ 36
Fashion	All Data	≈ 92.12	≈ 92.11	≈ 0.279	≈ 90.78	≈ 90.79	≈ 0.253	≈ 89.73	≈ 17
Fashion	**EnBaSe**	≈ 92.41	≈ 92.41	≈ 0.262	≈ 91.06	≈ 91.06	≈ 0.246	≈ 79.62	≈ 10
Fashion	Random	≈ 91.29	≈ 91.29	≈ 0.308	≈ 88.98	≈ 88.99	≈ 0.303	≈ 88.50	≈ 10
CIFAR-10	All Data	≈ 90.93	≈ 90.93	≈ 0.395	≈ 86.43	≈ 86.44	≈ 0.403	≈ 85.90	≈ 1,019
CIFAR-10	**EnBaSe**	≈ 89.22	≈ 89.21	≈ 0.441	≈ 82.05	≈ 82.03	≈ 0.558	≈ 78	≈ 501
CIFAR-10	Random	≈ 90.12	≈ 90.10	≈ 0.450	≈ 82.23	≈ 82.23	≈ 0.554	≈ 81.78	≈ 505
CIFAR-100	All Data	≈ 79.33	≈ 79.33	≈ 0.948	≈ 72.32	≈ 72.31	≈ 0.947	≈ 71.92	≈ 16,393
CIFAR-100	**EnBaSe**	≈ 77.38	≈ 77.37	≈ 1.114	≈ 67.68	≈ 67.69	≈ 1.134	≈ 65.04	≈ 10,398
CIFAR-100	Random	≈ 77.26	≈ 77.26	≈ 1.150	≈ 67.41	≈ 67.41	≈ 1.129	≈ 66.66	≈ 12,775

Note: The MNIST and Fashion models were trained for ten epochs, whereas the CIFAR-10 and CIFAR-100 models were trained for 50 epochs.

TABLE 10: Performance Comparison (iid) with Different Works.

Dataset	Architecture	Author	Training	Validation
Mnist	CNN-2	[70]	97.07	-
Mnist	CNN	[71]	98.54	97.76
Mnist	CNN-1	[70]	99.21	-
Mnist	**CNN**	**EnBaSe**	99.28	98.64
Fashion	**CNN**	**EnBaSe**	92.41	91.06
Fashion	MCNN-14	[72]	93.08	-
Fashion	CNN-1	[73]	95.22	88.95
Fashion	CNN-2	[73]	98.01	93.11
CIFAR-10	CNN-2	[74]	85.90	-
CIFAR-10	CNN-1	[74]	87.57	-
CIFAR-10	**CNN**	**EnBaSe**	89.22	82.05
CIFAR-10	CNN	[71]	98.91	97.71
CIFAR-100	FC-CNN-Lab	[75]	42.26	-
CIFAR-100	CNN-1	[76]	63.50	-
CIFAR-100	CNN-2	[76]	68.60	-
CIFAR-100	CNN	**EnBaSe**	77.38	67.68

Note: The values of accuracy provided by the authors for training and validation.

The experiments in this section demonstrate that the EnBaSe method may slightly underperform compared to training with the entire dataset, resulting in a minor accuracy reduction. This difference becomes more pronounced in complex benchmarks, such as CIFAR-100 (Table 9).

Furthermore, the experiments demonstrate that, although computationally efficient, reducing the cost and convergence time by 50%—the EnBaSe method exhibits an average accuracy loss of 3%. On CIFAR-100, this reduction is even more significant, achieving an accuracy of 67.68%, compared to 72.32% for the "All Data" method during validation.

The experiments on CIFAR-100, with only 500 images per class, challenged the EnBaSe method, which selects 250 high-quality images per class. Despite this careful selection, the limitation in sample size contributes to suboptimal results in this more complex scenario.

Finally, in Table 10, in the scenario with centralized and distributed data in a iid manner, EnBaSe performance demonstrates consistency with other studies available in the literature. The model achieves the highest accuracy on the MNIST dataset, with 99.28% in training and 98.64% in validation.

In the case of Fashion-MNIST, EnBaSe presents an accuracy of 92.41%, in contrast to one of the best models, which registers 98.91%. EnBaSe obtains 91.01% for validation, while the best model reaches 93.11%.

For CIFAR-10, EnBaSe achieved an accuracy of 89.22%, while the best model achieved 98.91%. EnBaSe achieved 82.05% in validation, compared to 97.71% for the best model.

In the case of CIFAR-100, EnBaSe presented an accuracy of 77.38%, surpassing the second-best model, which achieved 68.60%. In this scenario, it is observed that most of the authors did not provide validation samples, limiting more detailed comparisons.

E. EVALUATION IN A NON-IID SCENARIO

Table 11 lists the datasets evaluated for the FL architecture. The selected datasets have varying levels of complexity and represent diverse challenges in the field of ML. For instance, the MNIST dataset is a starting point for state-of-the-art experiments. In contrast, Fashion-MNIST offers greater complexity and poses a more significant challenge than MNIST. Finally, the CIFAR-10 and CIFAR-100 datasets present substantial challenges for the state-of-the-art models.

The MNIST, Fashion-MNIST, CIFAR-10, and CIFAR-100 datasets, the distribution skew discussed in Section V-C was applied. In this scenario, each node (e.g., device) participating in the training within the FL architecture receives data allocations randomly.

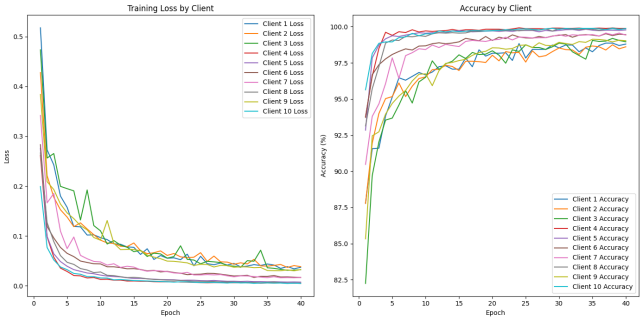
This randomness results in a non-uniform distribution, meaning that specific classes are not guaranteed to be equally represented or assigned to each node. This method aims to ensure that the differences observed between nodes arise from intrinsic randomness rather than systematic bias. This approach aims to replicate the diversity present in real-world IoT scenarios, allowing each node to contribute unevenly, thereby simulating realistic conditions.

TABLE 11: Datasets in FL architecture.

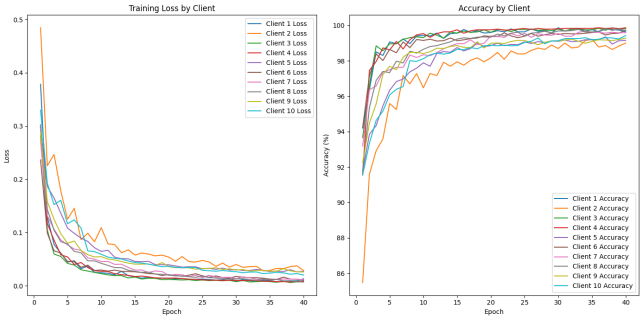
Dataset	Description	Classes	Distribution	Architectures
MNIST	Handwritten digits	10	non-iid	FL
Fashion-MNIST	Clothing items	10	non-iid	FL
CIFAR-10	Various objects	10	non-iid	FL
CIFAR-100	100 categories	100	non-iid	FL

In the non-iid scenario, we conducted 240 experiments to identify the model's behavior in a non-iid environment, focusing on optimizing data quality and reducing computational costs.

Figures 11 (a) and 11 (b) show the results of FL applied to the MNIST dataset, using the FedAvg and FedProx algorithms, respectively. Similarly, Figures 12 (a) and 12 (b) show the same algorithms applied to the Fashion-MNIST dataset. In addition, the results for the CIFAR-10 set are illustrated in Figures 13 (a) and 13 (b), while those for CIFAR-100 are represented in Figures 14 (a) and (b).



(a) FedAvg convergence.

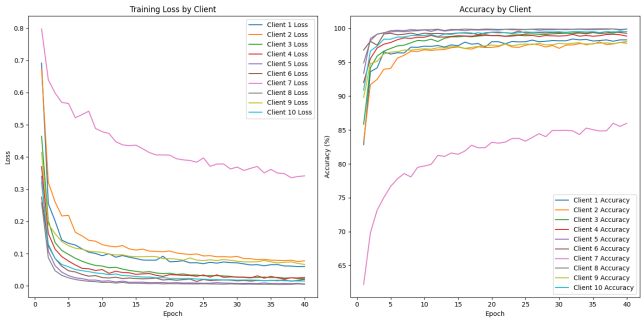


(b) FedProx convergence.

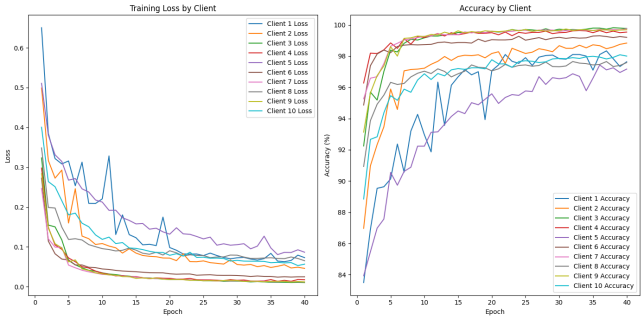
FIGURE 11: MNIST: EnBaSe (FedAvg & FedProx).

Table 12 presents the results obtained from 240 experiments, providing the average precision, recall, F1-score, accuracy, loss, and training time. These results enable the observation of how entropy influences data quality and reduces computational and energy costs.

Half of the dataset was randomly selected using the Random method to ensure a fair comparison and follow the same methodology. As a result, this allowed for comparing training with a complete dataset and training using an entropy-based selection. All models in the table followed the same training pattern using 50 epochs with ten available nodes participating.

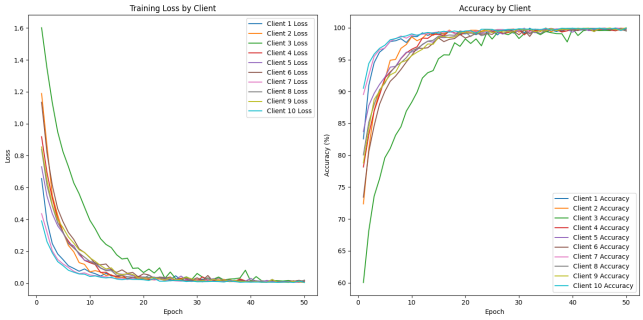


(a) FedAvg convergence.

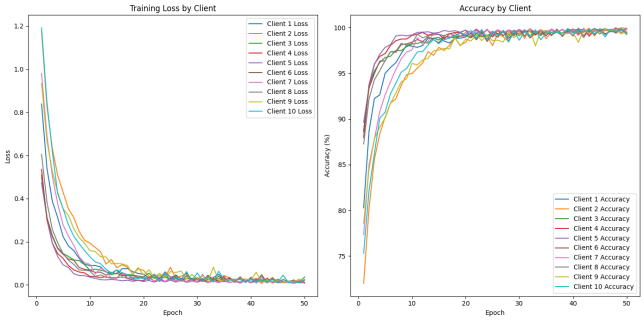


(b) FedProx convergence.

FIGURE 12: Fashion-MNIST's: EnBaSe (FedAvg & FedProx).



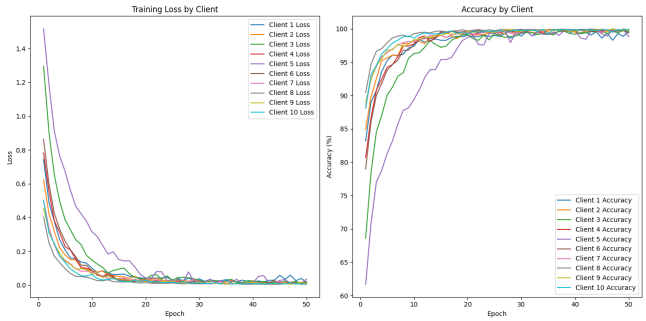
(a) FedAvg convergence.



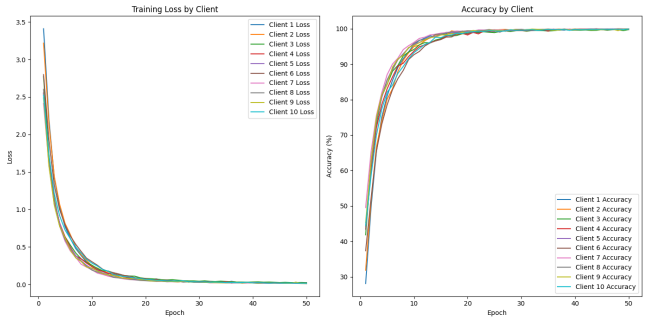
(b) FedProx convergence.

FIGURE 13: CIFAR-10: EnBaSe (FedAvg & FedProx).

Consequently, it is possible to highlight the quality selection strategy using EnBaSe, which demonstrates a sig-



(a) FedAvg convergence.



(b) FedProx convergence.

FIGURE 14: CIFAR-100: EnBaSe (FedAvg & FedProx).

nificant reduction in training time compared to the model using the entire dataset. This observation was confirmed by examining the training times of all the EnBaSe and Random datasets.

Additionally, we noted a slight decrease in the overall accuracy of the EnBaSe model compared to the entire dataset model. Finally, in the global aggregation method, FedProx EnBaSe proved particularly effective over time, significantly reducing computation time while maintaining acceptable accuracy levels.

The results presented in Table 12 compare EnBaSe and All Data using the same neural network architecture and hyperparameters. Although EnBaSe reduces the computational cost by approximately 50%, and the metric results are close and comparable, there are minor losses in precision and accuracy across nearly all models.

This slight performance drop is evident in most datasets, where EnBaSe demonstrates small reductions in metrics such as accuracy, precision, recall, and F1-score compared to the entire dataset (e.g., All Data) as shown in Table 12.

Moreover, Table 12 also indicates that, in some cases, the loss values are slightly higher than those observed with All Data, suggesting that EnBaSe has not yet achieved complete convergence. This observation highlights the importance of more refined adjustments to hyperparameters and the neural network architecture to improve performance.

F. BENCHMARK: MULTIPLE CLIENTS AND HIGH LOAD

This section revisits the experiment conducted under the non-iid scenario using the same criteria for the evaluation metrics specified in Section V-C. The neural networks employed were the same as those introduced in Section V-D, and the detailed experimental configurations are given

TABLE 12: Average Experiment results with FedAvg and FedProx.

Dataset	Algorithm	Model Selection	Precision	Recall	F1-Score	Accuracy (%)	Loss	Time (s)
Mnist	FedAvg	All Data	≈74.30%	≈74.40%	≈71.90%	≈85.70%	≈0.253	≈1,096
Mnist	**FedAvg**	**EnBaSe**	≈66.10%	≈66.50%	≈62.60%	≈78.91%	≈0.193	≈586
Mnist	FedAvg	Random	≈42.50%	≈40.20%	≈34.60%	≈56.80%	≈0.248	≈555
Mnist	FedProx	All Data	≈69.26%	≈71.40%	≈67.48%	≈81.73%	≈0.179	≈1,180
Mnist	**FedProx**	**EnBaSe**	≈69.28%	≈68.93%	≈66.01%	≈81.26%	≈0.264	≈629
Mnist	FedProx	Random	≈42.55%	≈42.97%	≈37.81%	≈59.13%	≈0.264	≈606
Fashion	FedAvg	All Data	≈56.80%	≈52.70%	≈47.50%	≈62.21%	≈1.753	≈1,084
Fashion	**FedAvg**	**EnBaSe**	≈53.80%	≈49.70%	≈45.40%	≈58.90%	≈1.572	≈556
Fashion	FedAvg	Random	≈27.80%	≈25.50%	≈20.70%	≈39.90%	≈3.505	≈574
Fashion	FedProx	All Data	≈60.30%	≈57.10%	≈52.30%	≈66.60%	≈1.390	≈1,125
Fashion	**FedProx**	**EnBaSe**	≈55.70%	≈52.50%	≈47.60%	≈60.80%	≈1.709	≈601
Fashion	FedProx	Random	≈26.50%	≈24.80%	≈18.80%	≈35.90%	≈3.856	≈632
CIFAR-10	FedAvg	All Data	≈47.80%	≈41.50%	≈34.40%	≈43.12%	≈1.749	≈14,958
CIFAR-10	**FedAvg**	**EnBaSe**	≈43.90%	≈38.40%	≈32.70%	≈40.32%	≈1.889	≈8,259
CIFAR-10	FedAvg	Random	≈8.96%	≈9.97%	≈2.72%	≈10.01%	≈4.274	≈8,205
CIFAR-10	FedProx	All Data	≈46.12%	≈38.75%	≈32.33%	≈39.53%	≈2.009	≈15,662
CIFAR-10	**FedProx**	**EnBaSe**	≈43.69%	≈37.39%	≈32.12%	≈40.42%	≈1.821	≈8,413
CIFAR-10	FedProx	Random	≈8.79%	≈9.99%	≈3.06%	≈10.09%	≈4.776	≈8,394
CIFAR-100	FedAvg	All Data	≈47.84%	≈41.18%	≈34.85%	≈43.24%	≈1.715	≈15,214
CIFAR-100	**FedAvg**	**EnBaSe**	≈44.70%	≈39.75%	≈33.98%	≈43.57%	≈1.744	≈8,131
CIFAR-100	FedAvg	Random	≈8.31%	≈9.96%	≈2.27%	≈10.00%	≈5.092	≈8,274
CIFAR-100	FedProx	All Data	≈55.41%	≈55.84%	≈51.94%	≈63.08%	≈1.834	≈18,020
CIFAR-100	**FedProx**	**EnBaSe**	≈48.01%	≈46.41%	≈42.99%	≈53.54%	≈2.198	≈8,202
CIFAR-100	FedProx	Random	≈27.70%	≈35.71%	≈28.29%	≈41.20%	≈3.345	≈8,418

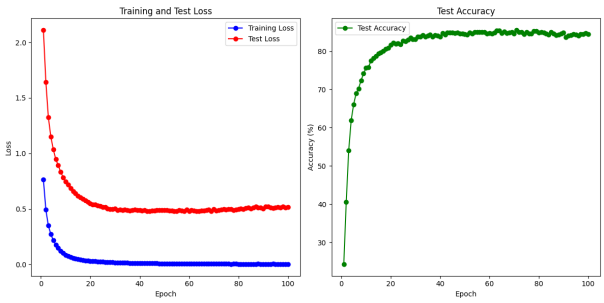
Note: The datasets were trained for 50 epochs with ten available clients (nodes).

in Section V-B. To better understand the general idea presented in the initial hypothesis of this work regarding the quality of local data at the edge (EnBaSe), we conducted a benchmark with FedProx, a model developed for challenges in more realistic scenarios, aligned with challenging skewed distributions.

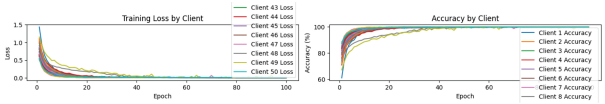
This test will help us understand how an edge-embedded algorithm for improving quality interacts with a more sophisticated global aggregation technique. In other words, we consider both sides of the federated learning process: the edge and the global neural network.

We kept the same settings for the neural network's architecture and parameters to test the model's efficacy under challenging conditions. We utilized the following hardware: 83 GB of RAM, 40 GB GPU, and 201 GB storage. The number of clients and epochs was increased to 50 and 100, respectively. We evaluated the algorithm's capacity using a large workload and data diversity.

This adjustment aims to replicate an advanced computational system for rigorous model analysis under high demand. In doing so, we can observe that with many clients (devices with datasets), the EnBaSe algorithm managed to approach smooth convergence.



(a) Global model convergence over 100 epochs for the CIFAR-10.

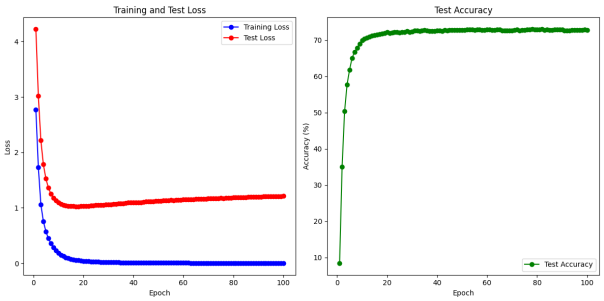


(b) The FedProx algorithm clients converged over 100 epochs for 50 clients at CIFAR-10

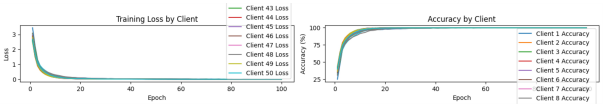
.

FIGURE 15: Global model and clients (nodes) convergence for 100 epochs and 50 clients at CIFAR-10 for the FedProx algorithm.

The results presented in Figures 15 and 16 for the **CIFAR-10** and **CIFAR-100** benchmarks, respectively, show the performance of the **FedProx** model using **EnBaSe** over 100 epochs with 50 clients (nodes). For CIFAR-10, the model achieved a precision of 82.20%, recall of 81.51%, F1-Score of 81.32%, and accuracy of 84.46%, with a loss value of 0.515, and a training time of 47,011.90 seconds. For CIFAR-100, the precision was 71.71%, recall 70.83%, F1-Score 70.69%, and accuracy 72.84%, with a loss value



(a) Global model convergence over 100 epochs for CIFAR-100.



(b) FedProx algorithm clients convergence over 100 epochs for 50 clients at CIFAR-100.

FIGURE 16: Global model and clients (nodes) convergence for 100 epochs and 50 clients at CIFAR-100 for the FedProx algorithm.

of 1.216 and a training time of 46,983.84 seconds.

Furthermore, the model reached the maximum neural network architecture accuracy plateau much earlier than the total number of training epochs. The EnBaSe model significantly reduced the processing time, approximately halving the computational cost. Therefore, this implies that the energy consumption and computational expenses would be considerably higher if the experiment were conducted again without the proposed EnBaSe algorithm.

Table 13 presents the benchmark test results of EnBaSe, along with a comparison with other recent studies using state-of-the-art methods to evaluate the effects of the EnBaSe algorithm on accuracy. Thus, we compiled optimal results from the experiments reported in this article.

TABLE 13: Performance of Models on CIFAR-10 and CIFAR-100 Datasets in the non-iid Scenario.

Dataset	Architecture	Author	Model	Acc (%)
CIFAR-10	CNN	[77]	Scaffold-Vanilla	26.65
	CNN	[77]	FedProx-Vanilla	27.42
	CNN	[77]	FedAvg-Vanilla	58.57
	CNN	[78]	AdaFedAdam	72.77
	CNN	[79]	FedCOME	75.88
	CNN	**Our Model**	**FedProx (EnBaSe)**	**84.46**
	CNN	[46]	FedPer++	85.09
	CNN	[47]	FedAvg (Adapted)	90.80
CIFAR-100	CNN	[77]	Scaffold-Vanilla	4.49
	CNN	[77]	FedProx-Vanilla	10.39
	CNN	[19]	FedHKD	29.88
	CNN	[79]	FedCOME	37.66
	CNN	[61]	Fed-IT	39.29
	CNN	[77]	FedAvg-Vanilla	40.36
	CNN	[80]	FedProx(FedFed)	70.02
	CNN	**Our Model**	**FedProx (EnBaSe)**	**72.84**

In this work, we compared the results of the proposed algorithm, obtained in Section VI-E (Table 12), where FL was applied, with the benchmark results presented in this

section (Table 13).

This comparison considered the increase in the number of connected devices from 10 to 50, as well as the increase in the number of training epochs from 50 to 100, representing a higher workload and an additional challenge for the algorithm (EnBaSe). The analysis focuses on the two most challenging datasets, CIFAR-10 and CIFAR-100, which provide a solid foundation for the evaluation.

We observed that, in the case of CIFAR-10 and CIFAR-100, the precision values for FedProx EnBaSe improved from 40.42% and 53.54% to 84.46% and 72.84%, respectively. These results indicate that the algorithm performs increasingly better as more devices are connected, even in more challenging scenarios.

Finally, Table 13 compares the results obtained with the state of the art. The most challenging datasets and the results obtained by different methods were considered. In the context of the CIFAR-10 dataset, in a non-iid scenario, the EnBaSe model achieved an accuracy of 84.46%, a performance comparable to that of FedPer++ (85.09%) and superior to that of FedCOME (75.88%).

In the case of the CIFAR-100 dataset, EnBaSe presented an accuracy of 72.84%, significantly outperforming other models, such as FedCOME (37.66%) and Fed-IT (39.29%). These results demonstrate the efficiency of EnBaSe in optimizing data selection, reducing computational costs without significantly compromising the accuracy of the models.

VII. DISCUSSION

We hypothesized that entropy can be used to measure data quality, considering that it quantifies the uncertainty or unpredictability in each node. A high entropy value indicates that data are highly unpredictable or vary within a node. As a result, each dataset in a node provides significantly different information, making it difficult to predict based on previous information.

Thus, the hypothesis was tested using the low entropy in each node, which was measured as a separation criterion to avoid asymmetry. For clarification purposes, this study analyzes hypotheses regarding data quality in nodes and the reduction of computational costs in the field of CV in collaboration with a laboratory involved in a project under contractual confidentiality, focusing on monitoring devices that complement data related to biosignals, where embedded systems have low processing power. This study, related to the field of CV, conducted initial experiments using a CNN, as it is relatively more straightforward and cost-effective.

We searched for keywords in conjunction with the names of the MNIST, Fashion-MNIST, CIFAR-10, and CIFAR-100 datasets, including time benchmark, training benchmark, training time comparison, training time performance, optimization training time, throughput training time, throughput training performance, and various combinations of these words. No studies in the literature have provided simulations of variations in the implemented hardware, training times, or times per epoch. Therefore, we compared the latest metrics from the most recent state-of-the-art models with those in this study.

One of the difficulties encountered in the literature during the development of this study was the need for more standardization of the experiments presented. Many studies provide only partial data without including essential metrics, such as validation, F1-score, recall, loss values, or the model's training time. Finally, one of the limitations of this study was the budget for the experiments, which led us to reduce the number of epochs to 50 and the number of nodes to 10 because of the large number of experiments conducted, as stated in Subsection VI-E.

Recent state-of-the-art approaches in FL architecture and centralized architectures predominantly focus on enhancing data homogeneity and addressing the heterogeneity and challenges associated with non-iid distributions.

However, such approaches often overlook fundamental considerations regarding the computational capacity of devices and the resources required for neural networks to achieve a high generalization capability.

This issue is critical for fostering more significant equity in the integrability and applicability of systems designed for architectures with low computational and energy capacities, which is the primary focus of this study.

Thus, the EnBaSe algorithm is efficient in optimizing the computational resources allocated to the neural network, such as the bandwidth used for global model updates in FL or data transfer in a centralized architecture, achieving up to a 50% reduction in communication.

This optimization also reduces the time required for training the neural network and, correspondingly, decreases the energy consumption, as models utilizing the EnBaSe algorithm complete training at least 50% faster.

To better understand this relationship. Energy is directly related to Consumption Power from devices and Execution time based on the following formula:

$$E = P(Whatts) \times Execution\_time(s)$$

Where E represents the energy consumed, P is the device's power in Watts, and $Execution\_time$ in seconds is the duration of use. Thus, a decrease in execution time directly impacts the battery saving for the IoT devices in the same percentage ratio as the reduction in execution time. This is an important benefit of the EnBaSe algorithm strategy.

Moreover, this advancement is particularly relevant for systems with energy and computational constraints, such as smart devices operating with limited resources, including drones or AI-powered medical equipment with low processing capabilities. The proposed approach applies to a wide range of systems.

As a result, the neural network applied to the dataset, referred to as "All Data," encompasses the entirety of the available data, whereas EnBaSe functions as a systematic

data mapping mechanism that effectively represents the map of an information system.

Finally, the analyses conducted in this study confirmed the feasibility of using an algorithm that enhances data quality while reducing computational and energy processing costs.

VIII. CONCLUSION

This study highlights the significance of assessing the quality of input data. Entropy experiments demonstrated that it retains probabilistic relevance even after linear transformations. Moreover, because entropy preserves consistent probabilistic values for the data and remains unaffected by linear geometric transformations, we can effectively leverage it for data selection before applying DA techniques and normalization, optimizing resource utilization. This approach is particularly beneficial for ML techniques, which are crucial in the context of the IoT, where the processing capacity of devices is often limited.

The primary method used in this study was implementing the EnBaSe algorithm, which selects data based on its entropy to improve data quality in iid and non-iid scenarios. The algorithm is designed to reduce computational costs while maintaining acceptable performance. It is instrumental in FL for IoT devices, where computational resources are limited, and non-iid mitigation is essential.

Additionally, the consistency of the results demonstrates that this approach has significant applications in real-world systems, where controlling data volume, data quality, and computational resources is critical. Examples include CV, image processing, traffic monitoring, automated inspection, digital health, diagnostics, and more.

The **main contributions of this study include**:

- A detailed study on the behavior of entropy in images and its distribution in CV;
- Analysis of the impact of linear transformations and normalization on data entropy;
- Reducing the computational cost of IoT edge devices;
- The presentation of detailed metrics such as accuracy, F1-score, recall, loss values, and model training time;
- Organizing the metrics to serve as a reference for future experiments;
- The comparison of accuracy in iid and non-iid scenarios with other experiments;
- A comprehensive literature review on data quality and FL; and
- The development of the EnBaSe algorithm, which efficiently selects high-quality data based on entropy analysis. This method reduces unnecessary computational processing, optimizes model convergence, and improves data selection for FL and centralized learning scenarios, particularly in resource-constrained environments such as IoT.

Currently, global aggregation models for FL are being developed to comprehensively address the challenges inherent in weight aggregation, including data distribution, customer selection, heterogeneity, and temporal and spatial dependencies, especially in scenarios characterized by non-iid samples. However, with these experiments, we have shown that it is possible to use algorithms in edge devices so that they can autonomously solve the challenges related to data distribution, simplifying the approaches present in the state-of-the-art.

Finally, one of the most important contributions of this study is an embedded algorithm that adaptively operates on edge IoT devices based on their distribution and seamlessly integrates with any aggregation model for FL.

In conclusion, the algorithm selects subsets of each class with lower entropy, aiming to increase the predictability of the system as discussed in Subsection II-C and Section IV. Thus, the inherent heterogeneity of each device is reduced based on its dataset, handling data asymmetry (e.g., extreme data, outliers) based on the distribution of each class, as specified in the algorithm IV-C. In this process, the sample tends to approximate a Gaussian, and the equivalent part with lower entropy is selected, as demonstrated in Subsection VI-B. Additionally, subsequent investigations will be carried out to elucidate the characteristic behaviors of extreme data when analyzed using various analytical techniques.

This section presents the main conclusions of this study. At the same time, Section IX explores the future directions for developing this research, discussing the key areas that warrant further investigation.

IX. FUTURE WORKS AND DIRECTIONS

Future research will focus on adapting the algorithm to more scenarios and integrating it with other optimization techniques. In summary, future research should expand the algorithm's applicability to other data types and FL scenarios.

Furthermore, related research on the convergence of global models aims to prevent excessive training and minimize the associated computing costs and energy consumption.

In future work, we intend to expand the experiments to create more significant variability in non-iid and iid scenarios, for example, by using the Dirichlet distribution, and increase comparisons by using other methods such as FedDyn, FedDF, Scaffold, and FedLAW. Thus, evaluating EnBaSe using aggregation methods that adopt different strategies allows us to identify the contexts in which performance remains more robust and where potential improvements are possible. This provides a deeper understanding of its behavior across diverse scenarios and solidifies its strengths and limitations for future advancements. Also, we will evaluate the individual impact of EnBaSe on energétic consumption on the IoT devices.

This simulated experiment indicated a strong tendency for adoption feasibility in the current AI algorithms. Since the experiments with EnBaSe demonstrated its initial feasibility, future work will conduct evaluations into experiments under

real-world operational conditions. In addition, we will also execute FL experiments on real-world IoT devices to capture fine-grained adjustments in our algorithm.

Finally, for future work, we intend to revisit the state of the art and explore other application domains of artificial intelligence for FL algorithms. In addition, we intend to investigate new metrics that complement or extend those currently used in the state-of-the-art.

ACKNOWLEDGEMENTS

This work was partially funded by Coordination for the Improvement of Higher Education Personnel - Brazil (CAPES) - Finance Code 001. A realização desta investigação foi parcialmente financiada profundos nacionais através da FCT - Fundação para a Ciência e Tecnologia, I.P. no âmbito dos projetos UIDB/04466/2020 e UIDP/04466/2020. This study was partially sponsored by Brazil's National Council for Scientific and Technological Development (CNPq) under grant no. 406517/2022-3. Partial support was provided via the CEREIA Project (# 2020/09706-7) São Paulo Research Foundation (FAPESP), FAPESP–MCTIC-CGI.BR in partnership with the Hapvida NotreDame Intermedica group.

REFERENCES

[1] J. Verbraeken, M. Wolting, J. Katzy, J. Kloppenburg, T. Verbelen, and J. S. Rellermeyer, "A survey on distributed machine learning," Acm computing surveys (csur), vol. 53, no. 2, pp. 1–33, 2020, DOI:10.1145/3377454.

[2] L. Caruccio, D. Desiato, G. Polese, and G. Tortora, "Gdpr compliant information confidentiality preservation in big data processing," IEEE Access, vol. 8, pp. 205 034–205 050, 2020, DOI:10.1109/ACCESS.2020.3036916.

[3] J. S. Yalli, M. H. Hasan, and A. Badawi, "Internet of things (iot): Origin, embedded technologies, smart applications and its growth in the last decade," IEEE Access, 2024, DOI:10.1109/ACCESS.2024.3418995.

[4] W. A. Al-Nbhany, A. T. Zahary, and A. A. Al-Shargabi, "Blockchain-iot healthcare applications and trends: a review," IEEE Access, 2024, DOI:10.1109/ACCESS.2023.3349187.

[5] J. B. Minani, F. Sabir, N. Moha, and Y.-G. Guéhéneuc, "A systematic review of iot systems testing: Objectives, approaches, tools, and challenges," IEEE Transactions on Software Engineering, 2024, DOI:10.1109/TSE.2024.3363611.

[6] S. Chintala, "Iot and ai synergy: Remote patient monitoring for improved healthcare," in 2024 4th International Conference on Innovative Practices in Technology and Management (ICIPTM). IEEE, 2024, pp. 1–6, DOI:10.1109/ICIPTM59628.2024.10563530.

[7] J. C. S. d. Anjos, K. J. Matteussi, F. C. Orlandi, J. L. V. Barbosa, J. S. Silva, L. F. Bittencourt, and C. F. R. Geyer, "A Survey on Collaborative Learning for Intelligent Autonomous Systems," ACM Comput. Surv., vol. 56, no. 4, p. 98:1–98:37, Nov. 2023, DOI:doi.org/10.1145/3625544.

[8] H. Sun, S. Li, F. R. Yu, Q. Qi, J. Wang, and J. Liao, "Toward communication-efficient federated learning in the internet of things with edge computing," IEEE Internet of Things Journal, vol. 7, no. 11, pp. 11 053–11 067, 2020, DOI:10.1109/JIOT.2020.2994596.

[9] L. U. Khan, W. Saad, Z. Han, E. Hossain, and C. S. Hong, "Federated learning for internet of things: Recent advances, taxonomy, and open challenges," IEEE Communications Surveys & Tutorials, vol. 23, no. 3, pp. 1759–1799, 2021, DOI:10.1109/COMST.2021.3090430.

[10] M. F. Criado, F. E. Casado, R. Iglesias, C. V. Regueiro, and S. Barro, "Non-iid data and continual learning processes in federated learning: A long road ahead," Information Fusion, vol. 88, pp. 263–280, 2022, DOI:doi.org/10.1016/j.inffus.2022.07.024.

[11] S. AbdulRahman, H. Tout, H. Ould-Slimane, A. Mourad, C. Talhi, and M. Guizani, "A survey on federated learning: The journey from centralized to distributed on-site learning and beyond," IEEE Internet of Things Journal, vol. 8, no. 7, pp. 5476–5497, 2020, DOI:10.1109/JIOT.2020.3030072.

[12] X. Ma, J. Zhu, Z. Lin, S. Chen, and Y. Qin, "A state-of-the-art survey on solving non-iid data in federated learning," Future Generation Computer Systems, vol. 135, pp. 244–258, 2022, DOI:doi.org/10.1016/j.future.2022.05.003.

[13] P. R. R. d. S. Junior, K. J. Matteussi, A. d. S. Veith, B. F. Zanchetta, V. R. Q. Leithardt, A. L. M., E. P. de Freitas, J. C. S. d. Anjos, and C. F. R. Geyer, "Boosting Big Data Streaming Applications in Clouds with BurstFlow," IEEE Access, vol. 8, p. 219124–219136, Dec. 2020, DOI:10.1109/ACCESS.2020.3042739.

[14] M. M. Bassiouni, R. K. Chakrabortty, K. M. Sallam, and O. K. Hussain, "Deep learning approaches to identify order status in a complex supply chain," Expert Systems with Applications, vol. 250, p. 123947, 2024, DOI:doi.org/10.1016/j.eswa.2024.123947.

[15] D. Rosendo, A. Costan, P. Valduriez, and G. Antoniu, "Distributed intelligence on the edge-to-cloud continuum: A systematic literature review," Journal of Parallel and Distributed Computing, vol. 166, pp. 71–94, 2022, DOI:doi.org/10.1016/j.jpdc.2022.04.004.

[16] J. Kang, Z. Xiong, D. Niyato, Y. Zou, Y. Zhang, and M. Guizani, "Reliable federated learning for mobile networks," IEEE Wireless Communications, vol. 27, no. 2, pp. 72–80, 2020, DOI:10.1109/MWC.001.1900119.

[17] A. Imteaj, U. Thakker, S. Wang, J. Li, and M. H. Amini, "A survey on federated learning for resource-constrained iot devices," IEEE Internet of Things Journal, vol. 9, no. 1, pp. 1–24, 2022, DOI:10.1109/JIOT.2021.3095077.

[18] S. Itahara, T. Nishio, Y. Koda, M. Morikura, and K. Yamamoto, "Distillation-based semi-supervised federated learning for communication-efficient collaborative training with non-iid private data," IEEE Transactions on Mobile Computing, vol. 22, no. 1, pp. 191–205, 2021, DOI:10.1109/TMC.2021.3070013.

[19] H. Chen, H. Vikalo et al., "The best of both worlds: Accurate global and personalized models through federated learning with data-free hyper-knowledge distillation," arXiv preprint arXiv:2301.08968, 2023, DOI:10.48550/arXiv.2301.08968.

[20] X. Yu, L. Li, X. He, S. Chen, L. Jiang et al., "Federated learning optimization algorithm for automatic weight optimal," Computational Intelligence and Neuroscience, vol. 2022, 2022, DOI:10.1155/2022/8342638.

[21] C. Tu, S. Zhao, and H. Deng, "Fedwns: Data distribution-wise node selection in federated learning via reinforcement learning," in 2023 26th International Conference on Computer Supported Cooperative Work in Design (CSCWD). IEEE, 2023, pp. 600–605, DOI:10.1109/CSCWD57460.2023.10152675.

[22] Z. Li, T. Lin, X. Shang, and C. Wu, "Revisiting weighted aggregation in federated learning with neural networks," in International Conference on Machine Learning. PMLR, 2023, pp. 19 767–19 788, url:https://proceedings.mlr.press/v202/li23s.html.

[23] S. Zheng, T. Ye, X. Li, and M. Gao, "Federated learning via consensus mechanism on heterogeneous data: A new perspective on convergence," arXiv preprint arXiv:2311.12358, 2023, DOI:10.1109/ICASSP48485.2024.10446892.

[24] Q. Sun, X. Li, J. Zhang, L. Xiong, W. Liu, J. Liu, Z. Qin, and K. Ren, "Shapleyfl: Robust federated learning based on shapley value," in Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2023, pp. 2096–2108, DOI:doi.org/10.1145/3580305.3599500.

[25] R. W. Condori Bustincio, A. M. de Souza, J. B. Da Costa, and L. Bittencourt, "Entropicfl: Efficient federated learning via data entropy and model divergence," in Proceedings of the IEEE/ACM 16th International Conference on Utility and Cloud Computing, 2023, pp. 1–6, DOI:doi.org/10.1145/3603166.3632611.

[26] Z. Zhang, Y. Zhang, D. Guo, L. Yao, and Z. Li, "Secfednids: Robust defense for poisoning attack against federated learning-based network intrusion detection system," Future Generation Computer Systems, vol. 134, pp. 154–169, 2022.

[27] N. Rodríguez-Barroso, E. Martínez-Cámara, M. V. Luzón, and F. Herrera, "Backdoor attacks-resilient aggregation based on robust filtering of outliers in federated learning for image classification," Knowledge-Based Systems, vol. 245, p. 108588, 2022.

[28] A. Al-Dulaimy, M. Jansen, B. Johansson, A. Trivedi, A. Iosup, M. Ashjaei, A. Galletta, D. Kimovski, R. Prodan, K. Tserpes et al., "The computing continuum: From iot to the cloud," Internet of Things, vol. 27, p. 101272, 2024.

[29] A. Ullah, T. Kiss, J. Kovács, F. Tusa, J. Deslauriers, H. Dagdeviren, R. Arjun, and H. Hamzeh, "Orchestration in the cloud-to-things compute continuum: taxonomy, survey and future directions," Journal of Cloud Computing, vol. 12, no. 1, pp. 1–29, 2023.

[30] A. Margara, G. Cugola, N. Felicioni, and S. Cilloni, "A model and survey of distributed data-intensive systems," ACM Computing Surveys, vol. 56, no. 1, pp. 1–69, 2023, DOI:doi.org/10.1145/3604801.

[31] Z. Lu, H. Pan, Y. Dai, X. Si, and Y. Zhang, "Federated learning with non-iid data: A survey," IEEE Internet of Things Journal, 2024, DOI:10.1109/JIOT.2024.3376548.

[32] J. Coutinho-Almeida, R. J. Cruz-Correia, and P. P. Rodrigues, "Evaluating distributed-learning on real-world obstetrics data: comparing distributed, centralized and local models," Scientific Reports, vol. 14, no. 1, p. 11128, 2024, DOI:doi.org/10.1038/s41598-024-61371-1.

[33] B. Rao, J. Zhang, D. Wu, C. Zhu, X. Sun, and B. Chen, "Privacy inference attack and defense in centralized and federated learning: A comprehensive survey," IEEE Transactions on Artificial Intelligence, 2024, DOI:10.1109/TAI.2024.3363670.

[34] L. Yuan, Z. Wang, L. Sun, S. Y. Philip, and C. G. Brinton, "Decentralized federated learning: A survey and perspective," IEEE Internet of Things Journal, 2024, DOI:10.1109/JIOT.2024.3407584.

[35] Y. Zhang, S. Gao, P. Cai, Z. Lei, and Y. Wang, "Information entropy-based differential evolution with extremely randomized trees and lightgbm for protein structural class prediction," Applied Soft Computing, vol. 136, p. 110064, 2023, DOI:doi.org/10.1016/j.asoc.2023.110064.

[36] S. Zhuge, Z. Zhou, W. Zhou, J. Wu, M. Deng, and M. Dai, "Noisy image segmentation utilizing entropy-adaptive fractional differential-driven active contours," Multimedia Tools and Applications, pp. 1–26, 2024, DOI:doi.org/10.1007/s11042-024-20058-5.

[37] M. Z. Hossain and A. Imteaj, "Fedavo: Improving communication efficiency in federated learning with african vultures optimizer," arXiv preprint arXiv:2305.01154, 2023, DOI:10.48550/arXiv.2305.01154.

[38] A. A. Al-Saedi, V. Boeva, and E. Casalicchio, "Fedco: Communication-efficient federated learning via clustering optimization," Future Internet, vol. 14, no. 12, p. 377, 2022, DOI:10.3390/fi14120377.

[39] F. C. Orlandi, J. C. S. Dos Anjos, J. F. d. P. Santana, V. R. Q. Leithardt, and C. F. R. Geyer, "Entropy to mitigate non-IID data problem on Federated Learning for the Edge Intelligence environment," IEEE Access, vol. 11, p. 78845–78857, Jul. 2023, DOI:10.1109/ACCESS.2023.3298704.

[40] Z. Tao, J. Wu, and Q. Li, "Preconditioned federated learning," arXiv preprint arXiv:2309.11378, 2023, DOI:10.48550/arXiv.2309.11378.

[41] B. Li, S. Chen, and K. Yu, "Feddkw–federated learning with dynamic kullback–leibler-divergence weight," ACM Transactions on Asian and Low-Resource Language Information Processing, 2023, DOI:10.1145/3594779.

[42] S. K. Lo, Y. Liu, Q. Lu, C. Wang, X. Xu, H. young Paik, and L. Zhu, "Toward trustworthy ai: Blockchain-based architecture design for accountability and fairness of federated learning systems," IEEE Internet of Things Journal, vol. 10, pp. 3276–3284, 2023, DOI:10.1109/JIOT.2022.3144450.

[43] Z. Li, H. Zhu, D. Zhong, C. Li, B. Wang, and Y. Yuan, "A novel framework for distributed and collaborative federated learning based on blockchain and smart contracts," in 2023 IEEE 3rd International Conference on Digital Twins and Parallel Intelligence (DTPI). IEEE, 2023, pp. 1–4, DOI:10.1109/DTPI59677.2023.10365414.

[44] W.-J. Yang and P.-C. Chung, "Significant weighted aggregation method for federated learning in non-iid environment," in 2023 Sixth International Symposium on Computer, Consumer and Control (IS3C). IEEE, 2023, pp. 330–333, DOI:10.1109/IS3C57901.2023.00095.

[45] K. M. M. Dolaat, A. Erbad, and M. Ibrar, "Enhancing global model accuracy: Federated learning for imbalanced medical image datasets," in 2023 International Symposium on Networks, Computers and Communications (ISNCC). IEEE, 2023, pp. 1–4, DOI:10.1109/ISNCC58260.2023.10323682.

[46] J. Xu, Y. Yan, and S.-L. Huang, "Fedper++: toward improved personalized federated learning on heterogeneous and imbalanced data," in 2022 International Joint Conference on Neural Networks (IJCNN). IEEE, 2022, pp. 01–08, DOI:10.1109/IJCNN55064.2022.9892585.

[47] S. Ullah and D.-H. Kim, "Federated learning convergence on iid features via optimized local model parameters," in 2022 IEEE International Conference on Big Data and Smart Computing (BigComp). IEEE, 2022, pp. 92–95, DOI:10.1109/BigComp54360.2022.00028.

[48] C. Huang, L. Xie, Y. Yang, W. Wang, B. Lin, and D. Cai, "Neural collapse inspired federated learning with non-iid data," arXiv preprint arXiv:2303.16066, 2023, DOI:10.48550/arXiv.2303.16066.

[49] Y. Qiao, H. Q. Le, and C. S. Hong, "Boosting federated learning convergence with prototype regularization," arXiv preprint arXiv:2307.10575, 2023, DOI:10.48550/arXiv.2307.10575.

[50] F. Sabah, Y. Chen, Z. Yang, M. Azam, N. Ahmad, and R. Sarwar, "Model optimization techniques in personalized federated learning: A survey," Expert Systems with Applications, p. 122874, 2023, DOI:10.1016/j.eswa.2023.122874.

[51] M. Ilić, M. Ivanović, V. Kurbalija, and A. Valachis, "Towards optimal learning: Investigating the impact of different model updating strategies in federated learning," Expert Systems with Applications, p. 123553, 2024, DOI:doi.org/10.1016/j.eswa.2023.122874.

[52] B. Li, Y. Wu, J. Song, R. Lu, T. Li, and L. Zhao, "Deepfed: Federated deep learning for intrusion detection in industrial cyber–physical systems," IEEE Transactions on Industrial Informatics, vol. 17, no. 8, pp. 5615–5624, 2020, DOI:10.1109/TII.2020.3023430.

[53] Z. Du, C. Wu, T. Yoshinaga, K.-L. A. Yau, Y. Ji, and J. Li, "Federated learning for vehicular internet of things: Recent advances and open issues," IEEE Open Journal of the Computer Society, vol. 1, pp. 45–61, 2020, DOI:10.1109/OJCS.2020.2992630.

[54] T. Gafni, N. Shlezinger, K. Cohen, Y. C. Eldar, and H. V. Poor, "Federated learning: A signal processing perspective," IEEE Signal Processing Magazine, vol. 39, no. 3, pp. 14–41, 2022, DOI:10.1109/MSP.2021.3125282.

[55] J. Wolfrath, N. Sreekumar, D. Kumar, Y. Wang, and A. Chandra, "Haccs: Heterogeneity-aware clustered client selection for accelerated federated learning," in 2022 IEEE International Parallel and Distributed Processing Symposium (IPDPS). IEEE, 2022, pp. 985–995, DOI:10.1109/IPDPS53621.2022.00100.

[56] Y. Li, X. Chao, and S. Ercisli, "Disturbed-entropy: a simple data quality assessment approach. ict express. 2022," DOI:doi.org/10.1016/j.icte.2022.01.006.

[57] H. Lee, "Towards convergence in federated learning via non-iid analysis in a distributed solar energy grid," Electronics, vol. 12, no. 7, p. 1580, 2023, DOI:10.3390/electronics12071580.

[58] C. Wu, Z. Li, F. Wang, and C. Wu, "Learning cautiously in federated learning with noisy and heterogeneous clients," in 2023 IEEE International Conference on Multimedia and Expo (ICME). IEEE, 2023, pp. 660–665, DOI:10.1109/ICME55011.2023.00119.

[59] V. N. Iyer, "A review on different techniques used to combat the non-iid and heterogeneous nature of data in fl," arXiv preprint arXiv:2401.00809, 2024, DOI:doi.org/10.1016/j.neucom.2021.07.098.

[60] L. Yan and X. Ge, "Entropy production-based energy efficiency optimization for wireless communication systems," IEEE Open Journal of the Communications Society, 2024, DOI:10.1109/OJCOMS.2024.3476454.

[61] S. M. Hamidi, R. Tan, L. Ye, and E.-H. Yang, "Fed-it: Addressing class imbalance in federated learning through an information-theoretic lens," in 2024 IEEE International Symposium on Information Theory (ISIT). IEEE, 2024, pp. 1848–1853, DOI:10.1109/ISIT57864.2024.10619204.

[62] J. Lu, H. Zhang, P. Zhou, X. Wang, C. Wang, and D. O. Wu, "Fedlaw: Value-aware federated learning with individual fairness and coalition stability," IEEE Transactions on Emerging Topics in Computational Intelligence, 2024, DOI:10.1109/TETCI.2024.3446458.

[63] R. S. Antunes, C. André da Costa, A. Küderle, I. A. Yari, and B. Eskofier, "Federated learning for healthcare: Systematic review and architecture proposal," ACM Transactions on Intelligent Systems and Technology (TIST), vol. 13, no. 4, pp. 1–23, 2022, DOI:doi.org/10.1145/3501813.

[64] C. Zhang, Y. Xie, H. Bai, B. Yu, W. Li, and Y. Gao, "A survey on federated learning," Knowledge-Based Systems, vol. 216, p. 106775, 2021, DOI:doi.org/10.1016/j.knosys.2021.106775.

[65] J. Wen, Z. Zhang, Y. Lan, Z. Cui, J. Cai, and W. Zhang, "A survey on federated learning: challenges and applications," International Journal of Machine Learning and Cybernetics, vol. 14, no. 2, pp. 513–535, 2023, DOI:doi.org/10.1007/s13042-022-01647-y.

[66] H. Azami, S. Sanei, and T. K. Rajji, "Ensemble entropy: A low bias approach for data analysis," Knowledge-Based Systems, vol. 256, p. 109876, 2022, DOI:doi.org/10.1016/j.knosys.2022.109876.

[67] S. Zheng, W. Yuan, X. Wang, and L. Duan, "Adaptive federated learning via new entropy approach," IEEE Transactions on Mobile Computing, 2024, DOI:10.1109/TMC.2024.3402080.

[68] A. Li, J. Sun, B. Wang, L. Duan, S. Li, Y. Chen, and H. Li, "Lotteryfl: Empower edge intelligence with personalized and communication-efficient federated learning," in 2021 IEEE/ACM Symposium on Edge Computing (SEC). IEEE, 2021, pp. 68–79, DOI:10.1145/3453142.3492909.

[69] S. X. Lee and G. J. McLachlan, "An overview of skew distributions in model-based clustering," Journal of Multivariate Analysis, vol. 188, p. 104853, 2022, DOI:doi.org/10.1016/j.jmva.2021.104853.

[70] P. Siddhartha et al., "Digit recognition of mnist handwritten using convolutional neural networks (cnn)," in 2023 International Conference on

Intelligent Systems for Communication, IoT and Security (ICISCoIS). IEEE, 2023, pp. 328–332, DOI:10.1109/ICISCoIS56541.2023.10100602.

[71] A. Tomar and H. Patidar, "Optimizing cnn model performance for mnist and cifar classification using rectified sigmoid and res activation functions," in 2023 7th International Conference On Computing, Communication, Control And Automation (ICCUBEA). IEEE, 2023, pp. 1–6, DOI:10.1109/ICCUBEA58933.2023.10392280.

[72] S. Saed, B. Teimourpour, K. Kalashi, and M. A. Soltanshahi, "An efficient multiple convolutional neural network model (mcnn-14) for fashion image classification," in 2024 10th International Conference on Web Research (ICWR). IEEE, 2024, pp. 13–21, DOI:10.1109/ICWR61162.2024.10533341.

[73] E. Xhaferra, E. Cina, and L. Toti, "Classification of standard fashion mnist dataset using deep learning based cnn algorithms," in 2022 International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), 2022, pp. 494–498.

[74] S. Aslam and A. B. Nassif, "Deep learning based cifar-10 classification," in 2023 Advances in Science and Engineering Technology International Conferences (ASET). IEEE, 2023, pp. 01–04, DOI:10.1109/ASET56582.2023.10180767.

[75] A. Sikdar, S. Udupa, and S. Sundaram, "Fully complex-valued deep learning model for visual perception," in ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2023, pp. 1–5, DOI:10.1109/ICASSP49357.2023.10095290.

[76] D. Li, Y.-C. Hsu, R. Sumikawa, A. Kosuge, M. Hamada, and T. Kuroda, "A 0.13 mj/prediction cifar-100 raster-scan-based wired-logic processor using non-linear neural network," in 2023 IEEE International Symposium on Circuits and Systems (ISCAS). IEEE, 2023, pp. 1–5, DOI:10.1109/ISCAS46773.2023.10181427.

[77] M. Morafah, M. Reisser, B. Lin, and C. Louizos, "Stable diffusion-based data augmentation for federated learning with non-iid data," arXiv preprint arXiv:2405.07925, 2024, DOI:doi.org/10.48550/arXiv.2405.07925.

[78] L. Ju, T. Zhang, S. Toor, and A. Hellander, "Accelerating fair federated learning: Adaptive federated adam," IEEE Transactions on Machine Learning in Communications and Networking, 2024, url:10.1109/TMLCN.2024.3423648.

[79] S. Zheng, T. Ye, X. Li, and M. Gao, "Federated learning via consensus mechanism on heterogeneous data: A new perspective on convergence," in ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2024, pp. 7595–7599, DOI:10.1109/ICASSP48485.2024.10446892.

[80] Z. Yang, Y. Zhang, Y. Zheng, X. Tian, H. Peng, T. Liu, and B. Han, "Fedfed: Feature distillation against data heterogeneity in federated learning," Advances in Neural Information Processing Systems, vol. 36, 2024, DOI:https://proceedings.neurips.cc/paper_files/paper/2023/hash/bdcdf38389d7fcefc73c4c3720217155-Abstract-Conference.html.

ERNESTO GURGEL VALENTE NETO is pursuing a master's degree in Teleinformatics Engineering at the Federal University of Ceará (UFC). He got a postgraduate degree in Data Science from Centro Universitário Farias Brito (FB UNI) in 2022, graduated in Computer Science and IT Management from the same institution in 2021. Additionally, he serves as a researcher at the Center of Reference in Artificial Intelligence (CRIA), focusing on AI, Data Science, and health in collaboration with the Processing for Data Analysis and Learning Systems Laboratory (SPIRAL).

SOLON ALVES PEIXOTO is currently pursuing a Ph.D. in Electrical Engineering. He completed a Doctorate in Teleinformatics Engineering at the Federal University of Ceará (UFC), a master's in Computer Science and a Bachelor's degree in Computer Engineering from the Federal Institute of Education, Science and Technology of Ceará (IFCE). He is a researcher at the Image Processing and Computational Simulation Laboratory (LAPISCO) at IFCE in Fortaleza, focusing on machine learning and pattern recognition in cloud environments. He has worked on embedded systems related to IoT and its integration, encompassing aspects of hardware, firmware, circuits, and electronics.

VALDERI REIS QUIETINHO LEITHARDT (Senior Member, IEEE) received the Ph.D. degree in computer science from INF-UFRGS, Brazil, in 2015. He is currently an Professor with the ISCTE - Instituto Universitario de Lisboa and a Researcher integrated with the Istar-Information Scienses, Technologies and Archicture Research Centre (ISTA), and Research Group Software Systems Enginnering. He is also a collaborating Researcher with the following research groups: COPELABS, Universidade Lusofona de Lisboa, Portugal, the Laboratory of Embedded and Distributed Systems, University of Vale do Itajai; (UNIVALI) Brazil, and Federal University of Ceara - UFC - Brazil, and the Expert Systems and Applications Laboratory, University of Salamanca, Spain. His research interests include distributed systems with a focus on data privacy, communication, and programming protocols, involving scenarios and applications for the Internet of Things, smart cities, big data and cloud computing.

JUAN FRANCISCO DE PAZ SANTANA received a degree in technical engineering in systems computer sciences, in 2003, and Engineering degree in computer sciences, a degree in statistics, and the Ph.D. degree in computer science from the University of Salamanca, Spain, in 2005, 2007, and 2010, respectively. He is currently a Full Professor at the University of Salamanca, where he is also a Researcher with the Expert Systems and Applications Laboratory (ESALab). He has been a co-author of published articles in several journals, workshops, and symposiums.

JULIO C. S. DOS ANJOS was awarded a Ph.D. degree in Computer Science by UFRGS (2017), as well as a Post-doc in Computer Science UFRGS (2021), and a Master's degree in Computer Science at UFRGS (2012) and a Bachelor's degree in Electrical Engineering at PUC/RS (1991). He has been a Professor at the Federal University of Ceara/Brazil in the Graduate Program in Teleinformatics Engineering (PPGETI/UFC), since 05/2022. His research interests: Distributed Systems, Hybrid infrastructures, Collaborative Learning, Intelligent Autonomous Systems, and Big Data Analytics with Deep Learning.

X. STATE-OF-THE-ART APPROACHES AND METHODS

In this section, we present a table of the state-of-the-art based on a systematic literature review. We identify the main approaches used, the scenarios in which the models have been applied, and/or the scenarios explored by the authors in their studies. We have also highlighted the main data sets and the metrics adopted and systematically identified the advantages and disadvantages pointed out by the authors in their respective works. In cases where this information was not explicitly mentioned, we analyzed possible positive or negative impacts based on the available evidence.

As we can see from the state-of-the-art analysis in Table 14, common approaches that employ customer aggregation, selection, sampling, or customer contribution measurement techniques stand out. This strategy is one of the most frequent and prevalent for dealing with device heterogeneity. Many of these methods are adaptive, seeking to extract characteristics to train neural networks and face the challenges posed by the data.

On the other hand, a second approach, which has been gaining relevance, is based on Information Theory. This line of research seeks to quantify information gain and system homogeneity based on entropy, as discussed by [18], [25], [35], [39], [56], [60], [61].

The main application scenarios in the literature predominantly involve the IoT or FL context, i.e., environments characterized by multiple connected devices and distributed data. While some authors direct their studies to scalability issues in multi-client systems, communication efficiency, or solving specific problems in the literature, others focus on specific application areas, such as medical or noisy data.

The primary datasets used are those widely known in the literature, such as MNIST, Fashion-MNIST, CIFAR-10, and CIFAR-100. These datasets represent different levels of difficulty and complexity in problems related to deep neural networks, with CIFAR-100 ultimately considered one of the most challenging benchmarks in the field.

On the other hand, some authors use other datasets for additional analysis, targeted at specific applications or designed for particular challenges. For example, in [56], a dataset of crop pest images was used, while in [35], proteins-related datasets were analyzed.

The primary metrics adopted often include accuracy, communication cost, efficiency, precision, recall, F1-Score, and Mean Absolute Error (MAE). In addition, some authors use metrics such as communication per round, TPR (True Positive Rate), TNR (True Negative Rate), entropy index, and Standard Deviation.

The main focus and advantages discussed in the papers are related to developing more adaptive systems capable of dealing with convergence, performance, efficiency, and difficulties associated with training neural networks in non-iid scenarios. These aspects are often pointed out as the main problems faced by FL in the literature.

Despite these advances, many solutions still present significant challenges in achieving global model convergence. These difficulties include loss of accuracy at the end of training, instability at critical moments of convergence, and disadvantages associated with adding complexity to architectures. Furthermore, this generally results in higher computational costs and problems related to increased modeling complexity.

A. DISCUSSION

We concluded that some articles address class balancing or heterogeneity, while few emphasize issues such as fairness in FL or bias mitigation. In addition, we found that most papers focus on communication efficiency and data heterogeneity, addressing FL by discussing the importance of privacy and the relevance of IoT devices.

Although some authors have explored strategies to reduce communication costs, improve convergence, energy efficiency, and gains in computational performance and generalization of neural networks, there is a lack of standardized metrics for FL experiments. Standardized metrics that consider the distribution of results and computational efficiency are particularly relevant in scenarios with IoT applications, where energy efficiency is as crucial as accuracy, especially on devices with limited computational resources.

Most studies mention FL training in IoT systems (e.g., smart cities, medical applications, industrial applications, etc.). However, these topics are not widely explored in problems where prediction needs to occur in real-time, such as continuous data flows. In addition to the adaptive models often used, it would be interesting to consider approaches based on online learning in the context of FL. These systems would need to adapt dynamically and make inferences with low latency, which would be highly relevant for smart cities.

Another aspect highlighted in the literature is that most domains of interest focus on image classification and analysis problems. However, other important domains, such as natural language processing, audio and video systems, and biosignal analysis, deserve more attention.

Finally, we conclude that many of the aspects analyzed in the current state of the art cover concerns related to privacy and the use of advanced technologies for data analysis. These studies offer a broader view of the existing gaps in the literature, allowing for the identification of future perspectives and challenges that are still open.

TABLE 14: State-of-the-Art Approaches, Metrics, and Scenarios.

Author	Date	Approach Used	Application Scenario	Datasets Used	Metrics Adopted	Advantages	Disadvantages
Itahara, Sohei, et al. [18]	2021	Aggregation with entropy reduction	IoT, FL and non-iid	MNIST, Fashion-MNIST, IMDb, Reuters	Accuracy, Communication Cost	Robustness against attacks and noise	Loss of Accuracy
Criado, Marcos F, et al. [10]	2022	Continual Learning	IoT, FL and non-iid	MNIST, SVHN, USPS, Office-31, Bing-Caltech256, COREL5000	Accuracy	Adapts to data distribution changes	Labeled data assumptions
Al-Saedi, et al. [38]	2022	Round-wise Clustering	Communication Efficiency	MNIST, Fashion MNIST, Fashion-MNIST, CIFAR-10	Accuracy	Reduces Communication, dynamic partitioning	Complex dynamic clusters
Yu, Xi, et al. [20]	2022	Dynamic Regularization	Multiple Clients	MNIST, Fashion-MNIST	Accuracy, Precision, Recall, AUC	Adaptive learning, resource-efficient	Instability near convergence
Ullah, Shan, et al. [47]	2022	Local Parameter Optimization	Image Classification	CIFAR-10	Accuracy	Improves efficiency and performance	Dependency on FedAVG
Xu, Jian, et al. [46]	2022	Custom Classifier	Heterogeneous Data	Fashion-MNIST, CIFAR-10	Accuracy	Handles data variations, reduces Communication	Risk of overfitting with insufficient local data
Li, Yang, et al. [56]	2022	Disturbed entropy	Agricultural pest recognition	Agricultural pest images	Accuracy	Reduces redundancy in datasets	Restricted to multi-class classification
Wolfrath, Joel, et al. [55]	2022	Client Clustering	Mobile, IoT Devices	Fashion-MNIST, CIFAR-10	Time-to-accuracy, model accuracy.	Faster convergence, efficient training	Dependence on stable distributions
Tu, Chengwu et al. [21]	2023	Node selection	IoT, FL and non-iid	MNIST, CIFAR-10	Accuracy, Communication Rounds	Fewer rounds, Global Accuracy	Poorly representative data
Li, Boyuan et al. [41]	2023	Knowledge-based dynamics	IoT, FL and non-iid	MNIST, Fashion-MNIST, CIFAR-10	Convergence, Accuracy	Better convergence and Accuracy	Complexity with diverse data
Yang, Wei-Jong et al. [44]	2023	Dynamic weights	IoT, FL and non-iid	MNIST, Fashion-MNIST, CIFAR-10	Global Accuracy, Communication Rounds	Less communication, dynamic adaptation	Additional processing for weights
Chen, Huancheng et al. [19]	2023	Feature extraction	IoT, FL and non-iid	CIFAR-10, CIFAR-100, SVHN	Local Accuracy, global Accuracy	Robustness for heterogeneous data	Dependency on hyper-knowledge
Zheng, Shu et al. [23]	2023	Client sampling	IoT, FL and non-iid	MNIST, Fashion-MNIST, CIFAR-10, CIFAR-100	Accuracy	Reflects global distribution	Quadratic computational complexity
Huang, Chenxi et al. [48]	2023	Global memory vectors	Heterogeneous devices	CIFAR-100	Weighted Accuracy, rounds-Accuracy	Reduces variance	Initial setup computations
Orlandi, Fernanda C. et al. [39]	2023	Entropy for Non-IID Data Mitigation	IoT Devices	MNIST, CIFAR-10	Accuracy, execution time	Mitigates impacts of non-IID data, lower power consumption	slight reduction in accuracy
Qiao, Yu et al. [49]	2023	Prototype regularization	Image classification	MNIST, Fashion-MNIST	Average Accuracy, Communication Efficiency	Fast convergence in non-iid scenarios	Prototype computation per round
Wu, Chenrui et al. [58]	2023	Prototypes and dynamic pseudo-labeling	Noisy data, imbalanced classes	CIFAR-10, CIFAR-100, Clothing1M	Accuracy, Precision, Recall	Stabilizes accurate pseudo-labeling	Computational cost
Sun, Qiheng et al. [24]	2023	Client contributions	Malicious clients, healthcare	CIFAR-10, Fashion-MNIST, Fed-ISIC2019	Accuracy, Convergence Rate	Robustness against poisoning attacks	Intensive computations, high computational cost
Milan Ilić et al. [51]	2023	Model updates	Medicine, IoT, FL and non-iid	Fashion-MNIST, LEAF, Adult Income, Body Signal of Smoking	F1-Score, Mean Absolute Error (MAE)	Flexible across multiple domains/tasks	Competing strategies with inferior performance
Condori Bustincio, et al. [25]	2023	Adaptive selection by entropy	Heterogeneity, communication overhead	CIFAR-10	Accuracy, Communication Cost	Reduces communication overhead	Limited generalization on datasets
Zhang, Yu, et al. [35]	2023	Differential evolution by entropy	Protein structure prediction	25PDB, FC699, D1189, D640	Accuracy, TPR, TNR, F1-Score	Robustness in feature selection	High complexity, requires fine-tuning
Hamidi, Shayan Mohajer, et al. [61]	2024	Entropy in loss function	Medical diagnosis	CIFAR-10, CIFAR-100, TinyImageNet	Accuracy, Standard Deviation	Better Accuracy on unbalanced datasets	Increased complexity
Yan, Litao, et al. [60]	2024	Entropy production model	Wireless communication	Physical and mathematical simulations	Entropy rate, Communication Cost	Optimized allocation, parallel processing	Complexity in modeling