Article Type: Feature

Pre-Deployment Complexity Estimation for Federated Perception Systems

KMA Solaiman, *ksolaima@umbc.edu, University of Maryland, Baltimore County, MD, USA*
Shafkat Islam, *islam59@pnw.edu, Purdue University Northwest, Hammond, IN, USA*
Ruy de Oliveira, *rdeolive@purdue.edu, Purdue University, West Lafayette, IN, USA*
Bharat Bhargava, *bbshail@purdue.edu, Purdue University, West Lafayette, IN, USA*

Abstract—Edge AI systems increasingly rely on federated learning to train perception models in distributed, privacy-preserving, and resource-constrained environments. Yet, before training begins, practitioners often lack practical tools to estimate how difficult a federated learning task will be in terms of achievable accuracy and communication cost. This paper presents a classifier-agnostic, pre-deployment framework for estimating learning complexity in federated perception systems by jointly modeling intrinsic properties of the data and characteristics of the distributed environment. The proposed complexity metric integrates dataset attributes such as dimensionality, sparsity, and heterogeneity with factors related to the composition of participating clients. Using federated learning as a representative distributed training setting, we examine how learning difficulty varies across different federated configurations. Experiments on multiple variants of the MNIST dataset show that the proposed metric strongly correlates with federated learning performance and the communication effort required to reach fixed accuracy targets. These findings suggest that complexity estimation can serve as a practical diagnostic tool for resource planning, dataset assessment, and feasibility evaluation in edge-deployed perception systems.

Introduction

Deploying perception models in edge environments introduces challenges that extend beyond model accuracy alone. Practical deployments must account for communication limitations, constrained computation, and fragmented data ownership, particularly when training data cannot be centralized. Distributed training paradigms have therefore become central to modern Edge AI systems, motivating the need for tools that help practitioners reason about training feasibility and cost before large-scale learning is attempted.

Federated learning (FL) has emerged as a widely adopted framework for distributed and privacy-preserving model training, especially in perception applications [1]. While a substantial body of research has focused on improving federated optimization strategies and convergence behavior [2], [3], considerably less attention has been paid to understanding why certain federated learning tasks are inherently more difficult than others. In practice, federated tasks trained using identical models and algorithms can exhibit markedly different convergence rates, communication costs, and final accuracy, differences that are often discovered only after training has already begun.

A key factor underlying this variability is the intrinsic structure of the data and the characteristics of the distributed environment. Prior work has shown that properties such as dimensionality, sparsity, and heterogeneity play a critical role in determining learning difficulty in perception domains [4], [5], [6]. However, most existing approaches assess complexity from the perspective of a single dataset or a centralized learning agent. Extending such analysis to distributed environments remains challenging, as complexity in these

XXXX-XXX © 2026 IEEE
Digital Object Identifier 10.1109/XXX.0000.0000000

settings depends not only on local data properties but also on how data are partitioned across multiple entities.

In this work, we adopt a measurement-oriented perspective and address the problem of *pre-deployment complexity estimation* for federated perception systems. Rather than proposing a new learning algorithm or federated optimization policy, our objective is to quantify how difficult a federated learning task is likely to be based solely on intrinsic data properties and distributed environment characteristics. This diagnostic view complements existing work on federated optimization by providing insight into task difficulty prior to training.

We propose a classifier-agnostic complexity estimation framework that captures three intrinsic attributes of perception data—dimensionality, sparsity, and heterogeneity—and integrates them with properties of the federated environment, such as the number and composition of participating clients. Federated learning is used as a representative distributed training paradigm, enabling the analysis of different federated configurations without modifying the underlying learning process. The resulting complexity metric allows different federated learning paths to be compared based on expected difficulty rather than observed performance.

To study the relationship between the proposed complexity metric and federated learning behavior, we conduct controlled experiments on multiple variants of the MNIST dataset in distributed settings. These experiments are designed to isolate complexity effects and examine how intrinsic and distributed complexity relate to federated accuracy and communication effort. The results demonstrate strong correlation between the proposed metric and both performance and training cost, suggesting that complexity estimation can serve as a practical diagnostic tool for edge-deployed perception systems.

Contributions

The main contributions of this work are as follows:

- We introduce an application-independent framework for estimating intrinsic complexity in perception datasets, based on dimensionality, sparsity, and heterogeneity, without reliance on classifier performance.
- We propose a unified complexity metric for federated learning environments that integrates intrinsic data properties with distributed environment characteristics, enabling comparison of federated training configurations.
- We empirically evaluate the proposed metric on multiple MNIST variants in distributed settings and demonstrate strong correlation with federated learning accuracy and communication effort.

Background and Related Work

Complexity Estimation in Perception Domains

Estimating the complexity of perception datasets and learning tasks has been studied from several perspectives, including information theory, uncertainty estimation [7], [8], [9], and dataset organization. Prior work has shown that intrinsic properties such as dimensionality, sparsity, and heterogeneity play an important role in determining learning difficulty in perception tasks.

Pereyda et al. [5] proposed a theoretical framework for measuring domain complexity and evaluated it using approximations derived from neural network models. Krusinga et al. [6] investigated dataset complexity by estimating probability densities of image distributions using generative adversarial networks, enabling the detection of outliers and domain shifts at the cost of significant computational overhead. Scheidegger et al. [4] examined classification difficulty using clustering-based metrics and lightweight neural probes, demonstrating that simple models can approximate dataset difficulty efficiently.

While these approaches provide valuable insight into dataset complexity, they are typically developed for centralized settings and often rely on classifier behavior or learned representations. In contrast, our work focuses on estimating intrinsic complexity directly from data properties and extends the analysis to distributed environments where data are partitioned across multiple entities.

Distributed and Federated Learning Context

Federated learning has emerged as a widely adopted framework for distributed and privacy-preserving model training [2]. A substantial body of research has focused on improving federated optimization and convergence behavior through methods such as FedAvg [2], FedAdaGrad [3] and FedYogi [3]. These approaches aim to improve training efficiency and robustness under heterogeneous data distributions and system constraints.

However, existing federated learning methods primarily address how to train a model efficiently given a fixed task and dataset distribution. They do not explicitly estimate the intrinsic difficulty of a federated learning task prior to training, nor do they model how data complexity and distributed environment characteristics jointly influence learning difficulty. As a result, task feasibility and communication cost are typically discovered only after training is underway.

Positioning of This Work

This work complements existing research by introducing a classifier-agnostic, measurement-oriented framework for estimating learning complexity in federated perception systems. By integrating intrinsic dataset properties with distributed environment characteristics, the proposed approach provides a pre-deployment diagnostic for assessing task difficulty. Federated learning is used as a representative distributed training setting, and the proposed metric does not modify or optimize federated learning algorithms. Instead, it aims to support feasibility assessment and resource-aware planning prior to training.

Complexity Estimation Framework

Overview

This section introduces the proposed complexity estimation framework for federated perception systems. The objective of the framework is to provide a *pre-deployment, classifier-agnostic diagnostic* for estimating how difficult a federated learning task is likely to be, based solely on intrinsic data properties and characteristics of the distributed training environment.

The framework decomposes learning difficulty into two complementary components. The first component captures the *intrinsic complexity of the data*, reflecting structural properties of the perception domain that influence learnability regardless of model architecture or optimization strategy. The second component captures the *complexity introduced by distribution*, reflecting how data are partitioned across multiple entities in a federated learning environment.

Intrinsic data complexity is characterized using three properties that have been shown to play a critical role in perception tasks: dimensionality, sparsity, and heterogeneity. Dimensionality captures the size and structure of the feature space, sparsity reflects redundancy and effective degrees of freedom in the data representation, and heterogeneity captures variability and diversity in the data distribution. These properties are computed directly from the data without reference to classifier performance or training dynamics.

Distributed complexity is modeled using federated learning as a representative training paradigm. Federated learning provides a concrete setting in which data are distributed across multiple participating entities, and learning proceeds through iterative communication rounds. Rather than modifying or optimizing the federated learning process, we use this setting to analyze how different federated configurations influence overall learning difficulty when combined with intrinsic data complexity.

The overall complexity of a federated learning task is expressed as a composite metric that integrates intrinsic data complexity with distributed environment characteristics. This formulation enables different federated training configurations to be compared in terms of expected difficulty, accuracy trends, and communication effort, without requiring model training to be executed in advance.

Figure 1 illustrates how different subsets of participating clients across communication rounds give rise to multiple federated learning paths, motivating the need to characterize task difficulty at the level of federated configurations rather than individual datasets. In the subsequent subsections, we formalize the intrinsic data complexity measures and define how they are integrated with federated environment characteristics to obtain the proposed federated complexity metric.

Intrinsic Perception Domain Complexity

Intrinsic complexity refers to properties of the data distribution itself that influence learning difficulty, independent of any specific model architecture, training algorithm, or optimization procedure. In perception tasks, these properties arise from the structure, variability, and organization of the data and can be assessed prior to training. In this work, intrinsic complexity is characterized along three complementary dimensions: dimensionality, sparsity, and heterogeneity.

Dimensionality Dimensionality captures the size and structure of the feature space that a learning system must represent. From a classifier-agnostic perspective, an upper bound on the dimensional complexity of a perception dataset can be defined using two factors: the number of samples and the number of features.

Let N_s denote the number of samples and N_d the number of feature dimensions. For supervised datasets, the number of output classes N_C is also included. An upper bound on environment dimensionality is given by

$$EC_{\text{upper}} = N_s \times N_d + N_C. \tag{1}$$

This upper bound reflects the maximum representational space implied by the dataset, regardless of the learning model. To refine this estimate, features

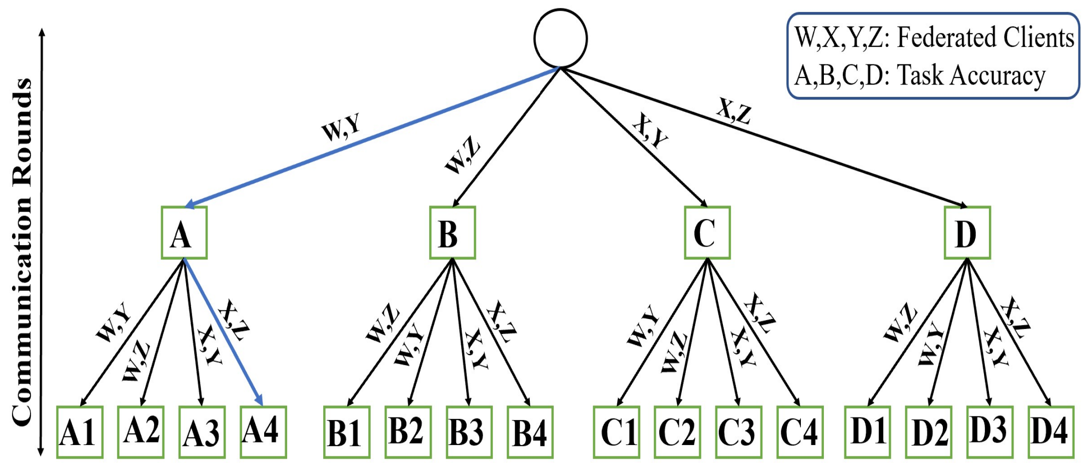

FIGURE 1. Illustration of federated learning paths across communication rounds. Each node represents an intermediate training state obtained from a subset of participating clients. Different paths correspond to distinct federated configurations, each associated with potentially different learning difficulty.

with zero variance across the dataset are removed, as they do not contribute to data variability. Let N_d^r denote the reduced number of features after removing zero-variance dimensions. The refined upper bound becomes

$$EC_{\text{upper}} = N_s \times N_d^r + N_C. \tag{2}$$

While upper-bound dimensionality reflects worst-case representational requirements, real-world perception data often lie on lower-dimensional manifolds. To capture this structure, we estimate intrinsic dimensionality (ID), defined as the minimum number of latent dimensions required to represent the data without information loss. Intrinsic dimensionality provides a lower-bound characterization of dimensional complexity that is sensitive to non-linear structure.

We estimate intrinsic dimensionality using the maximum likelihood estimator proposed in [10]. Given a dataset $\{x_i\}_{i=1}^n$, the estimator for intrinsic dimensionality m based on k nearest neighbors is defined as

$$m_k(x) = \left[\frac{1}{k-1}\sum_{j=1}^{k-1}\log\left(\frac{T_k(x)}{T_j(x)}\right)\right]^{-1}, \tag{3}$$

where $T_j(x)$ denotes the distance from sample x to its j-th nearest neighbor. This estimator provides a data-driven measure of latent dimensional structure without relying on class labels or classifier performance.

Sparsity Sparsity characterizes the degree to which informative structure in a dataset can be captured using a reduced set of components. In perception domains, sparsity reflects redundancy in feature representations and indicates how efficiently information can be compressed without significant loss.

We assess sparsity using unsupervised dimensionality reduction techniques that operate independently of class labels. Principal Component Analysis (PCA) [11] is used to quantify how much variance can be explained by a reduced number of linear components. A dataset that concentrates most of its variance in a small number of principal components is considered more sparse.

To capture non-linear structure, we additionally consider Isomap [12], a manifold learning method that preserves geodesic distances between samples. Isomap provides complementary insight into sparsity when data lie on curved or non-linear manifolds. Together, PCA and Isomap offer diagnostic indicators of how compactly a dataset can be represented, reflecting intrinsic sparsity in the perception domain.

Heterogeneity Heterogeneity measures the diversity and variability present within a dataset. In perception tasks, heterogeneity arises from variations in pixel intensities, textures, shapes, and other low-level features, and it directly influences learning difficulty.

We quantify heterogeneity using Shannon entropy [13], a standard information-theoretic measure of uncertainty. For grayscale images, entropy is defined as

$$H = -\sum_{i=0}^{n-1} p_i \log p_i, \tag{4}$$

where p_i denotes the probability of observing gray level i and n is the number of possible gray levels. Higher entropy values indicate greater diversity in pixel distributions and increased heterogeneity.

Taken together, dimensionality, sparsity, and heterogeneity provide complementary, classifier-agnostic indicators of intrinsic perception domain complexity. These measures are computed prior to training and serve as inputs to the federated complexity formulation described in the next subsection.

Federated Learning Complexity

Intrinsic data complexity captures the difficulty of learning from a single, local dataset, but it does not account for additional challenges introduced when data are distributed across multiple entities. In federated learning, data are partitioned across participating clients, and learning proceeds through iterative communication rounds. The distribution of data across entities and the patterns of participation during training can substantially influence learning behavior, even when the underlying model and optimization procedure remain fixed.

In this work, we use federated learning as a representative distributed training paradigm to model and analyze the impact of distribution on learning difficulty. Rather than modifying the federated learning process or proposing a new training strategy, we focus on characterizing how intrinsic data complexity interacts with distributed environment properties to influence task difficulty.

We represent a federated learning process as a conceptual tree, as illustrated in Figure 1, where each path corresponds to a distinct sequence of participating entities across communication rounds. Different paths represent different federated configurations and may lead to different learning outcomes due to variations in data exposure and aggregation order. This representation is used solely as an analytical tool to enumerate and compare federated configurations, not as a mechanism for controlling or selecting training behavior.

The overall complexity of a federated learning task is modeled as the combination of intrinsic data complexity and distributed environment complexity. We define the federated complexity metric as

$$F(d, X) = f(X) + f(d), \tag{5}$$

where $f(X)$ captures intrinsic data complexity, computed from dimensionality, sparsity, and heterogeneity measures, and $f(d)$ captures complexity arising from the distributed environment. Let $X = [x_1, x_2, \ldots, x_n]$ denote the vector of intrinsic complexity components.

The intrinsic complexity term is defined as

$$f(X) = \beta \|X\|_2 = \beta \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}, \tag{6}$$

where x_i denotes the intrinsic complexity of the local dataset associated with participating entity i, n is the number of entities involved in a given federated configuration, and β is a normalization constant.

To characterize the complexity introduced by data distribution across entities, we define the distributed environment complexity as

$$f(d) = \left(\frac{1}{m_1} + \frac{1}{m_2} + \cdots + \frac{1}{m_d} \right), \tag{7}$$

where d is the number of distinct entities in the federated environment and m_j denotes the frequency with which entity j appears across federated configurations. This formulation reflects the degree of fragmentation and imbalance in data distribution across entities, with higher values corresponding to more heterogeneous participation patterns.

Combining these components yields the federated learning complexity metric for a given federated configuration:

$$F(d, X) = \beta \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2} + \left(\frac{1}{m_1} + \frac{1}{m_2} + \cdots + \frac{1}{m_d} \right). \tag{8}$$

The proposed metric is used as a diagnostic indicator of learning difficulty. It enables comparison of different federated configurations in terms of expected accuracy trends and communication effort, without altering the learning algorithm or requiring training to be executed in advance. In the experimental section, we examine how this metric correlates with federated learning performance and effort across controlled distributed settings.

Experimental Setup and Evaluation

This section evaluates whether the proposed complexity framework provides a meaningful *diagnostic signal* for federated learning difficulty in perception domains. Specifically, we examine the relationship between the proposed complexity measures and two observable outcomes of federated learning: model accuracy and communication effort. The goal is not to optimize the federated learning process, but to assess whether complexity estimated *prior to training* correlates with learning behavior observed during training.

All experiments are conducted using Python 3.8 on a workstation equipped with an Intel Core i7 (8th generation) CPU and 16 GB of memory. Standard open-source libraries are used throughout to ensure reproducibility.

Datasets: We evaluate the proposed framework using three commonly used perception benchmarks: Handwritten-MNIST [14], Fashion-MNIST [15], and EMNIST-Digits [16]. These datasets provide controlled variations in visual complexity while maintaining identical input resolution and label structure, making them suitable for comparative complexity analysis.

Handwritten-MNIST and Fashion-MNIST each contain 70,000 grayscale images, with 60,000 training samples and 10,000 test samples. EMNIST-Digits contains 280,000 digit images. All images are of size 28×28 pixels with pixel intensities in the range $[0, 255]$.

Intrinsic Complexity Measurement: To characterize intrinsic dataset complexity, we compute the following measures for each dataset: (i) Shannon entropy as a measure of heterogeneity, (ii) sparsity estimated via principal component analysis (PCA), (iii) environment complexity based on feature variance thresholds, and (iv) intrinsic dimensionality (ID).

Entropy is computed using Scikit-image [17], PCA-based sparsity and environment complexity are computed using Scikit-learn [18], and intrinsic dimensionality is estimated using the Scikit-dimension package [19]. Table 1 summarizes the resulting measurements.

These values provide a relative ordering of dataset complexity that serves as input to the intrinsic complexity function $f(X)$ described in Section 3.

Federated Learning Configuration

Federated learning experiments are conducted using a fixed shallow convolutional neural network architecture across all settings. All clients employ identical model architectures and local training procedures. Model aggregation is performed using the FedAvg algorithm [2].

Unless otherwise stated, experiments use five federated clients with non-identical local data distributions. Each client performs a single local iteration per communication round, and all clients contain comparable amounts of data. For each configuration, training is run for 100 communication rounds. Federated test accuracy is evaluated on a held-out test set.

Evaluation Metrics

We evaluate federated learning behavior using two observable quantities: *accuracy* and *communication effort*. Accuracy is reported both as the maximum test accuracy achieved over all communication rounds and as the average accuracy across rounds. Communication effort is measured as the number of communication rounds required to reach a fixed accuracy threshold, set to 60% across all datasets.

To assess the diagnostic value of the proposed complexity metric, we analyze correlations between accuracy, effort, and the estimated complexity values. In particular, we examine how changes in federated environment complexity $f(d)$ and intrinsic complexity $f(X)$ relate to observed learning behavior, and whether the combined complexity metric $F(d, X)$ provides a stronger predictive signal than either component alone.

Results and Analysis

This section analyzes the relationship between the proposed complexity measures and observed federated learning behavior. The objective is to evaluate whether intrinsic dataset properties and distributed environment characteristics provide a reliable *diagnostic signal* for federated learning difficulty, as reflected in accuracy and communication effort. All results are reported using the experimental setup described in Section 4.

Intrinsic Complexity Across MNIST Variants

Table 1 summarizes the intrinsic complexity measurements for Handwritten-MNIST, EMNIST-Digits, and Fashion-MNIST. Across all metrics - heterogeneity, sparsity, environment complexity, and intrinsic dimensionality, Handwritten-MNIST consistently exhibits the lowest complexity, while Fashion-MNIST exhibits the highest. EMNIST-Digits occupies an intermediate position.

Although some metrics exhibit minor ordering differences (e.g., sparsity at higher variance thresholds), the overall trend is consistent. These results confirm that the selected intrinsic measures capture meaningful differences in dataset structure, even among datasets derived from the same base domain.

Federated Environment Complexity

Figures 2(b) and 2(c) illustrate the relationship between federated environment complexity $f(d)$ and federated learning accuracy for Handwritten-MNIST and Fashion-MNIST, respectively, with intrinsic complexity held fixed. As the number and diversity of participating clients increase, both maximum and average accuracy exhibit a decreasing trend.

Figure 3(a) further shows that communication effort increases monotonically with $f(d)$ for both datasets. These results indicate that distributed environment characteristics alone introduce measurable learning difficulty, independent of intrinsic dataset properties.

TABLE 1. Heterogeneity, Sparsity, Environment Complexity, and Intrinsic Dimensionality Measurement

Dataset	Heterogeneity	Sparsity ($r^2 = 80\%$)	Sparsity ($r^2 = 95\%$)	$EC_{upper}(v_\theta = 0)$	$EC_{upper}(v_\theta = 90)$	ID
Handwritten-MNIST	1.60	740	629	717	530	13.368
EMNIST-digits	2.86	751	685	697	557	14.095
Fashion-MNIST	4.11	760	594	784	745	14.547

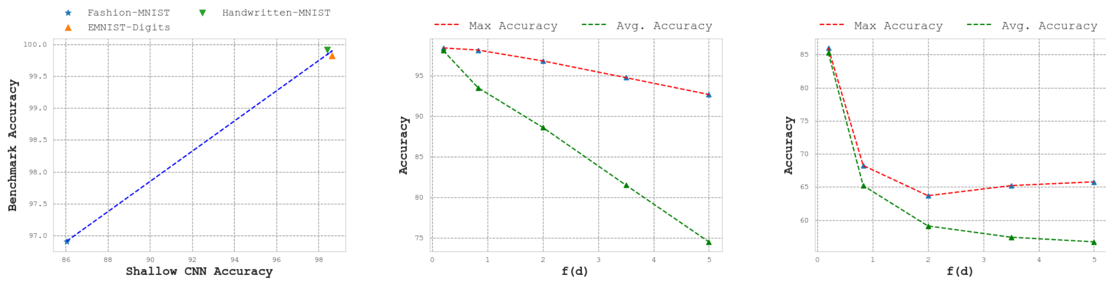

FIGURE 2. (a) Shallow CNN accuracy vs. benchmark accuracy for Fashion-MNIST, Handwritten-MNIST and EMNIST-Digits, (b) Federated environment complexity $f(d)$ vs. shallow federated learning accuracy for Handwritten-MNIST when $f(X)$ is fixed, (c) Federated environment complexity $f(d)$ vs. shallow federated learning accuracy for Fashion-MNIST when $f(X)$ is fixed.

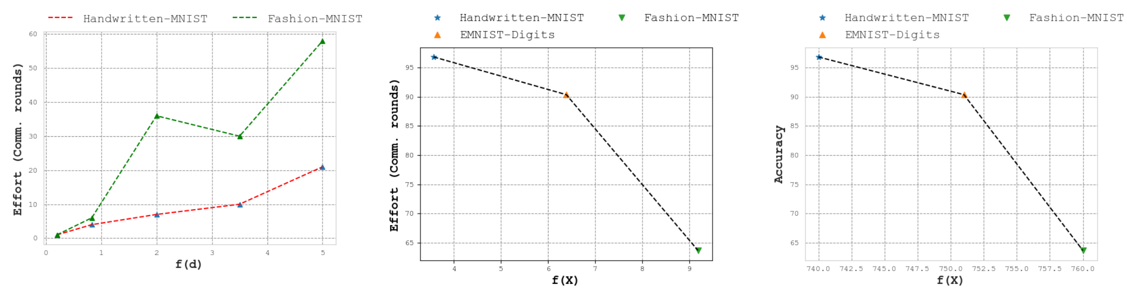

FIGURE 3. (a) Federated environment complexity $f(d)$ vs. effort (communication rounds) for MNIST and Fashion-MNIST, (b) Federated accuracy vs. Federated intrinsic function $f(X)$, reflecting heterogeneity (entropy), for Fashion-MNIST, Handwritten-MNIST and EMNIST-Digits, with $f(d)$ fixed at 2, (c) Federated accuracy vs. Federated intrinsic function $f(X)$, reflecting sparsity (number of sparse components for explaining 80% of variance), for Fashion-MNIST, Handwritten-MNIST and EMNIST-Digits with $f(d)$ fixed at 2.

Impact of Intrinsic Complexity on Federated Learning

Figures 3(b) and 3(c) examine the relationship between intrinsic complexity $f(X)$ and federated accuracy while holding the federated environment complexity constant. As intrinsic complexity increases—whether due to higher heterogeneity or increased sparsity—federated accuracy decreases.

Similarly, Figure 4(a) shows that communication effort increases as intrinsic complexity increases. These trends are consistent across all three MNIST variants, suggesting that intrinsic data properties directly influence learning difficulty in distributed settings.

Combined Complexity Metric

Figures 4(b) and 5 evaluate the relationship between the combined complexity metric $F(d, X)$ and federated learning accuracy. A strong negative correlation is observed between complexity and accuracy, with $R^2 = 0.81$ for maximum accuracy and $R^2 = 0.85$ for average accuracy.

Notably, average accuracy exhibits a slightly stronger correlation with the proposed complexity metric than maximum accuracy. This suggests that $F(d, X)$ better reflects sustained learning behavior rather than isolated peak performance, which is desirable for pre-deployment assessment.

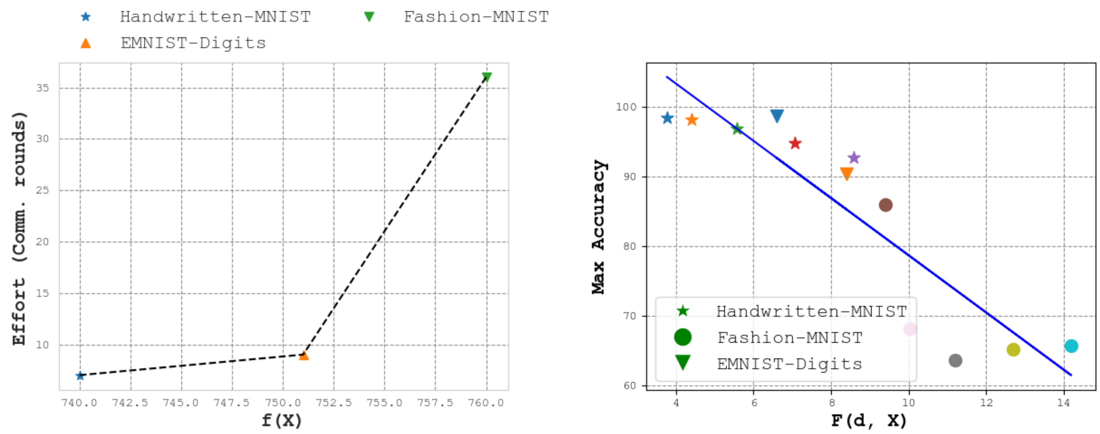

FIGURE 4. (a) Federated intrinsic function $f(X)$ (reflecting heterogeneity) vs. effort (communication rounds), (b) Federated complexity $F(d, X)$ vs. shallow Federated learning accuracy, considering maximum accuracy ($R^2 = 0.81$).

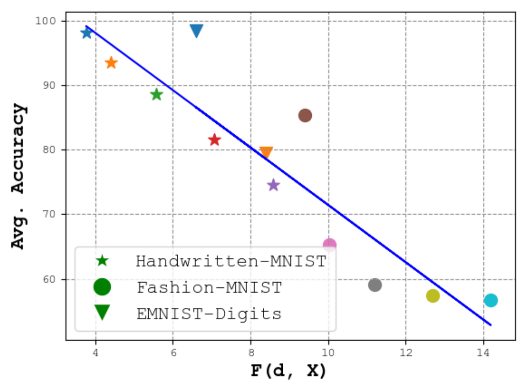

FIGURE 5. Federated complexity $F(d, X)$ vs. shallow Federated learning accuracy, considering average accuracy (R^2 = 0.85).

Summary of Findings

Across all experiments, intrinsic complexity, federated environment complexity, and their combination demonstrate consistent relationships with federated learning behavior. While neither intrinsic nor distributed complexity alone fully explains learning difficulty, their combination provides a stronger diagnostic signal.

These results support the use of complexity estimation as a complementary tool for understanding federated learning behavior prior to deployment, without modifying training algorithms or requiring extensive training runs.

Discussion

The results presented in this paper demonstrate that learning difficulty in federated perception systems is influenced by both intrinsic properties of the data and characteristics of the distributed training environment. By jointly modeling these factors, the proposed complexity metric provides a practical diagnostic signal that correlates with federated learning behavior, including achievable accuracy and communication effort.

A key implication of this work is that federated learning difficulty is not determined solely by the choice of model architecture or optimization algorithm. Even when training configurations are held constant, variations in dataset structure and client composition lead to measurable differences in convergence behavior and training cost. The proposed framework makes these differences explicit prior to training, enabling comparison of federated configurations without requiring extensive trial-and-error experimentation.

From an edge AI perspective, such pre-deployment complexity estimation is particularly valuable. Edge-deployed systems often operate under strict resource constraints, where communication budgets, energy consumption, and latency must be considered alongside accuracy. By providing an early indication of expected learning difficulty, the proposed metric can support feasibility assessment, dataset selection, and resource planning before committing to large-scale federated training.

It is important to emphasize that the proposed complexity metric is intended as a diagnostic tool rather than an optimization mechanism. The framework does not prescribe client selection policies, training schedules, or model adaptations, nor does it modify the federated learning process itself. Instead, it complements existing federated optimization methods by characterizing task difficulty independently of training dynamics. This separation allows complexity estimation to be applied broadly, regardless of the specific learning algorithm used.

The experimental evaluation is conducted using controlled perception benchmarks and simplified federated configurations. While these settings are sufficient to validate the diagnostic relationship between complexity and learning behavior, they do not capture the full diversity of real-world federated systems. In particular, factors such as extreme data imbalance, heterogeneous client hardware, partial participation, and more complex data modalities are not explored in this study. These limitations suggest natural directions for future work rather than shortcomings of the proposed framework.

Future extensions of this work may include evaluating complexity estimation on larger and more diverse perception datasets, incorporating additional sources of heterogeneity common in real-world federated deployments, and studying how diagnostic complexity measures interact with adaptive training strategies. Importantly, such extensions can build upon the current framework without altering its core objective: providing a principled, pre-deployment understanding of federated learning difficulty in edge AI systems.

Conclusion

This paper introduced a diagnostic framework for estimating learning complexity in federated perception systems prior to training. By combining intrinsic dataset properties—dimensionality, sparsity, and heterogeneity—with characteristics of the distributed training environment, the proposed metric provides a unified measure of federated learning difficulty that is independent of model architecture and optimization strategy.

Using federated learning as a representative distributed training paradigm, we empirically evaluated the relationship between the proposed complexity measures and observable learning behavior. Experiments on multiple MNIST variants demonstrated consistent correlations between estimated complexity, federated accuracy, and communication effort. In particular, the combined complexity metric showed strong correlation with both average and maximum federated accuracy, indicating its potential utility as a pre-deployment diagnostic indicator.

The proposed approach does not seek to optimize or control the federated learning process. Instead, it complements existing federated learning methods by providing insight into task difficulty before training begins. Such early-stage complexity estimation can support feasibility assessment, dataset comparison, and resource planning in edge-deployed perception systems, where training costs and communication constraints are critical considerations.

Future work will explore the applicability of the proposed framework to more diverse and large-scale perception datasets, as well as federated settings with greater heterogeneity in data volume, client availability, and system resources. These extensions aim to further assess the generality of complexity estimation as a diagnostic tool for federated learning in real-world edge AI environments.

REFERENCES

1. T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, “Federated learning: Challenges, methods, and future directions,” *IEEE Signal Processing Magazine*, vol. 37, no. 3, pp. 50–60, 2020.
2. H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, “Communication-efficient learning of deep networks from decentralized data,” in *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2017. [Online]. Available: http://arxiv.org/abs/1602.05629
3. S. Reddi, Z. Charles, M. Zaheer, Z. Garrett, K. Rush, J. Konečnỳ, S. Kumar, and H. B. McMahan, “Adaptive federated optimization,” *arXiv preprint arXiv:2003.00295*, 2020.
4. F. Scheidegger, R. Istrate, G. Mariani, L. Benini, C. Bekas, and C. Malossi, “Efficient image dataset classification difficulty estimation for predicting deep-learning accuracy,” *Vis. Comput.*, vol. 37, no. 6, pp. 1593–1610, Jun. 2021.
5. C. Pereyda and L. B. Holder, “Measuring the complexity of domains used to evaluate ai systems,” *ArXiv*, vol. abs/2010.01985, 2020.
6. R. Krusinga, S. Shah, M. Zwicker, T. Goldstein, and D. W. Jacobs, “Understanding the (un)interpretability of natural image distributions using generative models,” *ArXiv*, vol. abs/1901.01499, 2019.
7. M. Batty, R. Morphet, P. Masucci, and K. Stanilov, “Entropy, complexity, and spatial information,” *Journal of Geographical Systems*, vol. 16, no. 4, pp. 363–385, Sep. 2014. [Online]. Available: https://doi.org/10.1007/s10109-014-0202-2
8. P. Langley, “Open-world learning for radically autonomous agents,” *Proc. Conf. AAAI Artif. Intell.*, vol. 34, no. 09, pp. 13 539–13 543, Apr. 2020.
9. R. P. Cardoso, E. Hart, D. B. Kurka, and J. V. Pitt, “Using novelty search to explicitly create diversity

in ensembles of classifiers," in *Proceedings of the Genetic and Evolutionary Computation Conference*. New York, NY, USA: ACM, Jun. 2021.
10. E. Levina and P. Bickel, "Maximum likelihood estimation of intrinsic dimension," *Advances in neural information processing systems*, vol. 17, 2004.
11. I. T. Jolliffe and J. Cadima, "Principal component analysis: a review and recent developments," *Philosophical transactions of the royal society A: Mathematical, Physical and Engineering Sciences*, vol. 374, no. 2065, p. 20150202, 2016.
12. J. B. Tenenbaum, V. d. Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *science*, vol. 290, no. 5500, pp. 2319–2323, 2000.
13. C. E. Shannon, "A mathematical theory of communication," *ACM SIGMOBILE mobile computing and communications review*, vol. 5, no. 1, pp. 3–55, 2001.
14. L. Deng, "The mnist database of handwritten digit images for machine learning research," *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 141–142, 2012.
15. H. Xiao, K. Rasul, and R. Vollgraf, "Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms," 2017, cite arxiv:1708.07747Comment: Dataset is freely available at https://github.com/zalandoresearch/fashion-mnist Benchmark is available at http://fashion-mnist.s3-website.eu-central-1.amazonaws.com/. [Online]. Available: http://arxiv.org/abs/1708.07747
16. G. Cohen, S. Afshar, J. Tapson, and A. van Schaik, "Emnist: an extension of mnist to handwritten letters," 2017. [Online]. Available: https://arxiv.org/abs/1702.05373
17. S. van der Walt, J. L. Schönberger, J. Nunez-Iglesias, F. Boulogne, J. D. Warner, N. Yager, E. Gouillart, T. Yu, and the scikit-image contributors, "scikit-image: Image processing in python," *Peerj*, 2014.
18. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
19. J. Bac, E. M. Mirkes, A. N. Gorban, I. Tyukin, and A. Zinovyev, "Scikit-dimension: a python package for intrinsic dimension estimation," *Entropy*, vol. 23, no. 10, p. 1368, 2021.