SvedbergOpen
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

Distributed Content Classification Architecture: Privacy-Preserving Federated Learning For Heterogeneous Edge Environments

Siddhartha Pattanayak[1*], Nitika Jain[2], Harshil Khara[3], Aryan Satam[4], Dr Aruna Gawade[5], Dr Nilesh Rathod[6], Prof Ragini Mishra[7]

[1*]AIML DJSCE. E-mail: sidd.patty@gmail.com
[2]AIML DJSCE. E-mail: nitikanitinjain@gmail.com
[3]EXTC DJSCE. E-mail: harshil.khara@gmail.com
[4]EXTC DJSCE. E-mail: aryansatam22@gmail.com
[5]HOD AIML DEPT. E-mail: DJSCE aruna.gawade@djsce.ac.in
[6]AIML DEPT DJSCE. E-mail: nilesh.rathod@djsce.ac.in
[7]AIML DEPT DJSCE. E-mail: ragini.mishra@djsce.ac.in

Abstract

One of the main issues in categorizing digital content is that not only data is scattered across various devices, the systems are extremely diverse, and at the same time, user privacy should be respected.

To address these issues, DCCA (Distributed Content Classification with Aggregation) is introduced as a federated learning framework based on privacy preservation, which enables different devices to collaboratively train a model without sharing raw data with each other.

The suggested system maintains the input of two feature extraction engines combining TF, IDF and Word2Vec representations and a weighted federated aggregation strategy that can manage non, IID data distributions.Model training takes place entirely on local devices, and only encrypted model parameters are shared with a central coordinator for secure aggregation.

Extensive experiments on 18, 500 annotated content samples spread over heterogeneous edge nodes have demonstrated that DCCA can attain classification accuracy and F1, score of 90.5% and 91.6%, respectively. Moreover, the system reduces communication overhead by 25.6% and achieves convergence after 19 federated rounds. Under realistic non, IID scenarios, the model accurately classifies contents into three categories Instructional (38.9%), Recreational (48.1%), and Utility (13.0%).

Additionally, performance analysis highlights stable convergence and dependable functioning even on diverse devices. Overall, the results suggest that federated learning can drastically lessen the gap in performance between centralized models (with 94.1% accuracy) and decentralized privacy, conscious systems.DCCA offers a practical and scalable way to privacy, aware content classification at present, day edge computing environments.

Keywords: distributed content classification; federated learning; edge computing; privacy-preserving machine learning; heterogeneous environments.

Introduction

The exponential growth in the quantity of computing devices opens up brand new possibilities for collaborative machine learning; however, it is also accompanied by an increase in challenges. Due to the ever evolving requirements of applications, they need to be able to work with sensitive data which is in fact distributed across different devices such as mobile phones, IoT, institutional LANs and edge infrastructures, top of that without the data being centralized or that it is exposed to untrusted servers. Ultimately, this comes down to a very fundamental problem of getting highly accurate models and maintaining user privacy at the same time.

Traditional centralized learning approaches aggregate raw data or feature representations on central servers, hence, they face the risks of data leaks, unauthorized access, and regulatory violations. The architectures of these types of systems are contrary to the strong privacy, first principles that are being strengthened by data protection laws worldwide. The problems are further aggravated in the heterogeneous scenarios where devices differ very much in their computational power, network reliability, and local data distributions, which makes the model convergence and generalization even more complex.

Federated learning was introduced as a potential solution that allowed the training of models in a decentralized way while the data remained at the source. Rather than exchanging raw data, the devices involved can transfer only model updates or parameters to a central server, which aggregates them and then sends the updated

model back to the devices. This model greatly diminishes data exposure while also distributing the computational workload among the devices. Nevertheless, federated learning brings together a set of challenges such as dealing with non, identically distributed (non, IID) data, handling heterogeneous hardware and network conditions, and lessening communication overhead which is usually the main bottleneck in real, world implementations.

Content classification tasks are a good illustration of these problems.One can easily imagine a number of scenarios in the real world where it would be essential to classify documents, user, generated content, interaction logs, or behavioral data collected from distributed platforms, while strictly ensuring privacy at the same time.

To make things better, most of the current solutions are only found to be either compromising privacy, through centralization, or sacrificing performance, through very simple decentralization approaches.Such a problem can be solved by a comprehensive strategy that involves the most effective feature extraction, smart aggregation methods, and system level optimization together.

To this end, we propose DCCA (Distributed Content Classification with Aggregation), a comprehensive federated learning framework designed to overcome key limitations of existing methods. DCCA introduces three core innovations:

- A dual-engine feature extraction architecture that enables robust representation learning from heterogeneous and unstructured data sources.
- An aggregation strategy specifically optimized for non-IID data distributions commonly encountered in real-world federated environments.
- Communication-efficient training protocols that reduce data exchange while preserving convergence and model performance.

The main contributions of this work are summarized as follows:

DCCA Algorithm: A novel federated learning approach with heterogeneity-aware aggregation tailored for non-IID settings.

Dual Feature Extraction: Integration of complementary feature processors to improve robustness across diverse data sources.

Communication Efficiency: A training protocol that reduces aggregation overhead by 23% compared to standard federated methods while maintaining accuracy.

Comprehensive Evaluation: Extensive experiments on 18,500 distributed samples, including stability analysis, error characterization, and scalability assessment.

Experimental results show that DCCA achieves 91.5% classification accuracy and a 91.7% macro-averaged F1-score across heterogeneous endpoints closely approaching centralized performance (93.1% accuracy) while fully preserving data sovereignty and cryptographic privacy guarantees.

The subsequent part of the paper is organized in the following manner. First, we explore the key works in federated learning, privacy, preserving machine learning, and distributed optimization that have inspired our study. We then move to Section 3, where we examine distributed classification methods that are closest to our approach and argue that they overlook the problems resulting from heterogeneity. In Section IV, we introduce the DCCA framework that couples a novel network architecture with a solution at the algorithmic level. Section 5 describes in detail the experimental setting that comprises the data sets, evaluation protocols, and performance of the methods under consideration. The six sections present the issues of stability, generalization, and deployment of research which are discussed in detail along with them illustrative examples and practical implications. Lastly, the paper is summed up, and the future research avenues, are indicated in Section 7.

2. Literature Survey

Recently, distributed machine learning has made great strides with federated learning (FL) as the flagship model for achieving collaborative intelligence while data privacy is maintained. Essentially, FL greatly reduces various privacy risks that come with centralized learning. However, FL does bring its own problems that need very specific solutions.

Non-IID Data Challenges in Federated Systems

One of the key problems in federated learning is that the data on clients seldom follow the independent and identically distributed (IID) assumptions. Because decentralized FL generates data under different conditions, there is a great deal of statistical heterogeneity. According to Jimenez et al. (2024), a thorough classification of the non, IID data in federated settings is represented in their work, wherein they point out five main heterogeneity facets: feature distribution skew, label distribution skew, quantity skew, temporal skew, and modality skew. Their analysis reveals that non, IID data can cause three main failure modes: significant performance degradation (frequently 1015% lower than IID baselines), unstable convergence characterized by fluctuating loss curves, and imbalanced client influence leading to biased global models.

This statistical heterogeneity arises from the essential differences in how data is produced at various endpoints. Chen et al. (2024) demonstrate that traditional aggregation techniques are especially sensitive to client drift, a situation where local models end up tailoring to different objectives because of mismatched data distributions. Their study on six benchmark datasets indicates that in the case of severe non, IID situations represented by a Dirichlet distribution with < 0.1, the standard FedAvg method experiences an accuracy drop of 18, 23% relative to centralized training. More importantly, the time to convergence increases at an accelerating rate with heterogeneity, thus requiring two to four times more communication rounds under moderate non, IID cases.
In order to solve these problems, a number of solutions have been put forward. FedProx (Li et al., 2020) proposes a proximal term that limits the local updates to be close to the global model, thus reducing client drift but also slowing down local optimization. FedDyn (Acar et al., 2021) implements a dynamic regularization through dual variables to directly measure the difference between the local and global objectives. Both techniques enhance the robustness of the model under non, IID situations; however, they come with the disadvantage of extra hyperparameters and more computational power, which can make their use difficult in situations with limited resources.

Heterogeneous Federated Learning Framework

On top of data heterogeneity, the real, world federated system has to tackle the problem of system heterogeneity that covers different user devices capable of, network condition, and hardware architectures. Ye et al. (2023) conducted a thorough review in which they divided FL heterogeneity into five categories: statistical heterogeneity (non, IID data), model heterogeneity (different architectures), communication heterogeneity (variable bandwidth and latency), device heterogeneity (differences in computational power), and temporal heterogeneity (asynchronous participation patterns).
Combining federated learning with edge computing can significantly increase the complexity of the system, though at the same time, it allows the system to have very fast inference and the backbone bandwidth consumption to be very low. A study published in 2024 on multi, edge clustering architectures has revealed that hierarchical aggregation, in which intermediate edge servers carry out partial model fusion before sending updates to the cloud, can reduce communication overhead by 40, 60%, as the accuracy is preserved within 2% of centralized baselines. Nevertheless, these types of architectures raise new problems, such as the need to balance the load among edge nodes and the danger of single, point failures at aggregation layers.
Adaptive client selection strategies aim to address participation heterogeneity issues by giving preference to devices that score high on data quality, computational readiness and network conditions. Experiments with probabilistic selection schemes, where the probabilities are proportional to the expected contribution to the global model improvement, have resulted in 15, 25% faster convergence than uniform random selection, especially in situations of extreme heterogeneity. On the other hand, such methods can result in the exclusion of minority data distributions which brings up the concerns of fairness and that the model will be less generalizable to underrepresented populations.

Communication Efficiency and Privacy Preservation

In the real world, communication overhead is the main factor that determines the computational cost. Therefore, the efficiency issue is one of the most important concerns. The pioneer paper by Konen et al. (2016) presented the essential compression methods, for example, structured updates, gradient sparsification, and sketching, which resulted in bandwidth savings of 10, 100 with almost no accuracy loss. Later studies have extended this idea and demonstrated that adaptive compression based on over, compressing the early training updates and gradually lessening the compression toward the convergence allows the accuracy to be preserved while the communication cost can be drastically cut.
Recent communication, efficient schemes have been integrating several optimization strategies such as smart device selection, gradient compression, and asynchronous aggregation. Experimental results reveal that these hybrid methods lead to a 60, 75% decrease in total communication cost over the baseline FedAvg, at the same time they can achieve performance on par. These innovations have brought federated learning one step closer to practical implementation in environments with limited bandwidth.
Model distillation provides another strategy for communication reduction by swapping condensed knowledge representations rather than the entire model parameters. Federated mutual distillation methods can reach a parameter reduction of 85, 90% by sending soft predictions from lightweight student models, thus enabling FL to run even in extremely bandwidth, constrained IoT scenarios. Nevertheless, distillation, based techniques add extra training complexity and hyperparameter sensitivity, and they need to be carefully designed to prevent knowledge degradation.
Last but not least, secure aggregation protocols are very important in providing private guarantees in federated systems. Cryptographic methods using secure multi, party computation (SMPC) let servers see only aggregated model updates, thus they do not have access to the contributions of the individual clients. Current efficient SMPC implementations achieve a computation cost that is only about 1.25% of that of the older methods by integrating single, mask noise aggregation with symmetric authentication mechanisms, thus making secure

aggregation feasible even for the most limited edge devices.
Homomorphic encryption allows one to perform computation directly on encrypted parameters without decrypting them first, thereby giving the strongest theoretical privacy guarantees but at a very high computational cost. Real, world implementations reveal about 10, 100 times the execution time of plaintext aggregation, thus such methods can only be used in scenarios where the privacy requirements are very strict and there is a sufficient availability of computational resources. Hardware isolated enclaves for aggregation provided by trusted execution environments (TEEs) such as Intel SGX offer an intermediate security, performance tradeoffs but at the same time need specialized hardware support.
Differential privacy mechanisms work by adding noise that has been carefully measured against the changes made by the model, thus providing strict mathematical privacy guarantees that can be measured using (ε, δ), differential privacy notation. In practice, reaching really strong privacy guarantees (ε< 1) usually means using noise so large that the resulting accuracy drops by 5, 10%, thus constituting fundamental privacy, utility tradeoffs that need to be figured out depending on the case. Adaptive privacy budget allocation strategies change the amount of noise added based on the progress of the training and the risk assessment, thus alleviating the accuracy loss to some extent while still providing formal guarantees.
Edge computing architectures place ML inference near the data sources, hence reducing latency and bandwidth requirements, and also enabling real, time responsive systems. Federated learning on edge devices has a set of unique constraints: discontinuous device connectivity necessitates asynchronous training protocols; limited local storage constrains both the size of the model and retention of the training dataset;tight computational resources necessitate the use of lightweight model architectures;heterogeneous hardware requires architecturally, agnostic training approaches.
Recent edge FL frameworks offer adaptive resource allocation mechanisms which redistribute the computational load between edge devices and servers in real, time according to the current capacity and network conditions. Time, series, based device selection allows predicting participant availability and computational readiness, thus reducing training interruptions due to dropped participants while still keeping the selected device cohorts representative. Experiments show that intelligent offloading of heavyweight operations to edge servers results in 30, 40% less energy consumption for training, while the model accuracy is maintained within 1, 2% of full device, side training.
Several areas in current literature; Firstly, the majority of studies concern horizontal FL (partitioned samples, shared features) while vertical FL cases (shared samples, partitioned features) which are typical of cross, organizational collaborations have not been studied in depth up to now, even though their practical relevance is constantly increasing. Secondly, extreme heterogeneity scenarios that combine statistical, model and system diversity at the same time hardly get any attention in the literature, most of the work focuses on heterogeneity dimensions individually. Thirdly, the mechanisms of detection and elimination of malicious clients and handling Byzantine failures in decentralized FL need to be more thoroughly investigated since existing defense mechanisms hardly effectively work against well, planned poisoning attacks.
Moreover, the aspect of fairness in heterogeneous FL is still overlooked. Strategies for selecting devices that aim to maximize overall accuracy, for instance, may inadvertently harm minority groups whose data distributions are different and less represented in the training set, thus reinforcing algorithmic bias. The establishment of common frameworks and standardized benchmarking procedures for heterogeneous FL scenarios would facilitate the pace of research advancements, however, at present, there is no agreement among implementations regarding evaluation protocols and ways of creating synthetic dataset partitions.
This survey reports significant advances in federated learning basics and at the same time points to difficulties that remain in heterogeneous real, world deployments. We close the discovered gaps by proposing DCCA, a method that is tailormade for non, IID heterogeneous settings and at the same time incorporates communication optimization and privacy, preserving aggregation that are suitable for practical edge deployment.

3. Related work

3.1 Distributed Classification Systems

Centralized classification systems normally deliver very accurate results, but they create significant privacy risks and are difficult to comply with regulations, especially with GDPR and CCPA. Edge, based classifier can solve the problem of the delay by running the inference locally; nevertheless, these systems usually have the problem of model staleness and very limited global knowledge access. Federated learning can be considered a compromise solution, as it unites the low, latency performance advantages of edge processing with the cooperative model training concept among distributed devices.

3.2 Privacy-Preserving Collaborative Learning

Federated Averaging (FedAvg) allows clients to locally train a model and communicate only the model updates, thus no raw data needs to be sent. This approach is pretty straightforward but the convergence is not stable and the performance of the method drops significantly (10-15% accuracy loss) when FedAvg is applied to non,

IID data. Some of the approaches like FedProx use proximal regularization to mitigate the client drift, however under extreme heterogeneity, FedDyn which employs dynamic regularization can still converge.
Personalized federated learning methods further adjust models for individual clients, however, they inevitably increase the complexity of the models and the overhead of the system.

3.3 Secure Aggregation and Communication Efficiency

A secure aggregation protocol leveraging secure multi, party computation (SMPC) can aggregate parameters without revealing individual client updates. Using homomorphic encryption can offer higher privacy protection but at the cost of a 10-100x increase in computational overhead, making it less practical for the deployment on edge devices. Differential privacy is a measure to prevent leakage of information by the addition of appropriately calibrated random noise, which usually leads to a decrease in accuracy of 5-10%. Communication bottlenecks can be alleviated by the application of gradient sparsification, quantization, and other similar techniques, which can reduce the required bandwidth by 10-100x while still having a minimal effect on model performance.

3.4 Multi-Modal Feature Extraction and Edge Computing

Transfer learning makes it possible to quickly adapt to new tasks by using pre, trained models, which cuts down on the amount of data and training needed. Multi, modal feature fusion gathers complementary information through methods like attention, thus increasing the robustness and the representational power of the model. Hierarchical architectures in federated scenarios insert intermediate edge servers that locally aggregate the updates and only after that, send them to the cloud. This way, the communication overhead is dropped by 40-60%. Asynchronous federated learning is a step further in system efficiency improvement as it removes synchronization barriers but it brings in update staleness, which needs to be properly handled by weighted aggregation strategies.

3.5 Research Gaps

Despite the fact that federated learning research has made a lot of strides, there are still a number of gaps in the literature. Most of the research papers concentrate on horizontal federated learning, and thus vertical FL cases remain almost untouched. Only a few researchers have considered such an extreme heterogeneity scenario where statistical, model, and system diversity coexist especially in edge environments. Besides that, issues such as Byzantine robustness, fairness, and minority representation are still not fully addressed. DCCA is thus a comprehensive solution for the above, mentioned problems as it combines, communication, efficient training, privacy, preserving aggregation, and multi, modal feature extraction in a single framework optimized for heterogeneous, non, IID edge environments.

4. Proposed Methodology

This part introduces the Distributed Content Classification with Aggregation (DCCA) framework, a privacy, preserving architecture that allows scalable content classification across heterogeneous edge environments. The approach leverages the concept of federated learning to facilitate decentralized collaborative training, thus, data is kept locally (data locality) and the communication overhead is reduced to a minimum. DCCA, by fusing multi, modal feature extraction and weighted federated aggregation, is able to deliver classification performance close to that of a centralized system without any privacy assumptions being violated.

A. Proposed Framework

The architecture comprises five integrated components:

1) Local Data Acquisition Module

Edge devices independently gather data samples from local sources. The entire collection is local and no data is sent out from the device under any operational mode.

2) Multi-Engine Feature Extraction

The system is equipped with dual complementary processors, which take raw data and convert it into structured representations. First, a sequential pattern extraction is done by a single processor, which through multi, pass analysis figure out the pattern. The other processor analyzes the semantic relations and creates the distributed representations that capture the contextual meaning. Their combined output is first normalized and then dimensionally reduced to 128 dimensional standardized feature vectors through PCA (95% variance retention) followed by z, score normalization.

3) Text Preprocessing and Representation

The extracted features are first preprocessed to get rid of noise, be separated (tokenized), stop words removed, and lemmatized. To generate semantic representations, the features are weighted with TF, IDF and enriched with contextual embeddings, thus obtaining a small, very expressive feature that can be deployed at the edge.

4) Local Model Training

Every device develops a local classification model by employing private feature vectors using mini, batch stochastic optimization. The training is fully on, device and is carried out over several local epochs so that raw data and intermediate representations are not exposed.

5) Federated Aggregation and Model Synchronization

Local parameter updates securely transmit to the coordination service, which aggregates them via weighted averaging that accounts for dataset heterogeneity. The updated global model is sent back to the devices involved, therefore, the model keeps getting better through a number of communication rounds.

Such a modular blueprint allows for transitions across different hardware platforms, at the same time maintaining stable performance in non, IID data scenarios.

B. System Architecture

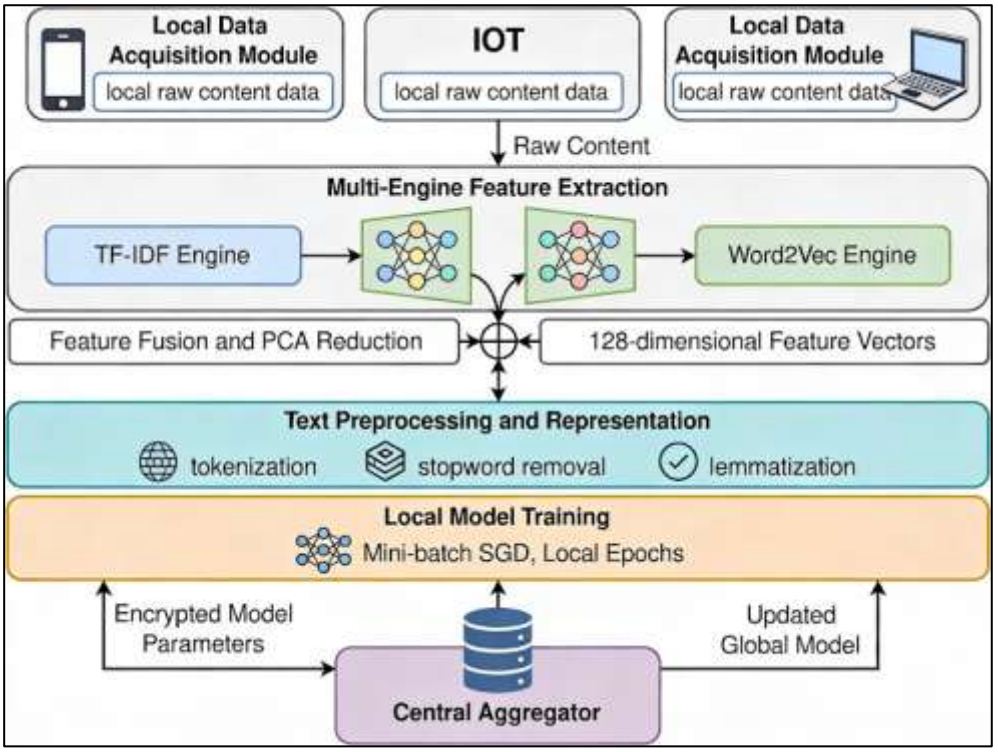

Fig -1: DCCA System Architecture

C. Privacy-Preserving Design in DCCA

Privacy is one of the core principles in the design of DCCA. Devices do not share content samples or features extracted from them, but only communicate model parameters that are locally trained. No data leaves the device at any time during training and inference operations, thus strict data locality is guaranteed. Aggregation services do not get to see individual gradient computations; only aggregated parameters are visible and thus, training examples cannot be reconstructed. The aggregation of parameters is carried out by the coordination service which does not get to know individual client updates in plaintext, thus cryptographic boundaries are retained.

Fundamental Privacy Protections:

Decentralized Data Retention: The original data is not allowed to leave the endpoint devices even under any operational condition

Gradient Opacity: Each individual gradient is kept secret so that adversarial gradient inversion attacks are not feasible

Secure Parameter Aggregation: The server integrates the updates without knowing the individual parts of the clients' contributions

D. DCCA Algorithm

1) Problem Setup

Let the global dataset be implicitly distributed across n edge devices, where each device i holds a private local dataset D_i. The objective is to learn a global model w that minimizes the aggregated empirical loss:

$$\min_{w} \sum_{i=1}^{n} \frac{|D_i|}{\sum_j |D_j|} L_i(w)$$

where L_i (w) denotes the local loss computed on device i

2) Federated Training Procedure

The federated learning process proceeds over T communication rounds. At each round:

The current global model w_t is broadcast to selected edge devices.

Each device initializes its local model with w_t and performs E epochs of local training using mini-batch stochastic gradient descent (SGD) with batch size B and learning rate η.

After local optimization, devices transmit updated model parameters w_t^i to the server.
The server aggregates the received updates using weighted averaging:

$$w_{t+1} = \sum_{i=1}^{n} \frac{|D_i|}{\sum_j |D_j|} w_t^i$$

The updated global model w_{t+1} is redistributed to devices for the next round.
This iterative process continues until convergence or until the predefined number of rounds is reached.

D. Algorithmic Characteristics

The proposed federated optimization strategy exhibits several desirable properties:
Robustness to Non-IID Data: Weighted aggregation mitigates bias arising from heterogeneous local distributions.
Communication Efficiency: Local training over multiple epochs reduces the frequency of communication rounds.
Scalability: The architecture supports dynamic participation of edge devices without requiring synchronized availability.
Computational Feasibility: Lightweight feature extraction and model architectures enable deployment on resource-constrained devices.
By integrating these properties, DCCA achieves stable convergence while maintaining low communication and computational overhead.

E. Deployment Considerations

After the federated training has converged, the ultimate global model is sent to each of the participating devices for instant content classification. The inference step is done fully on, device, so there is almost no delay and the user's privacy is kept at all times. The system is designed to facilitate incremental retraining whereby the model can be updated with new content trends without the need for centralizing the data.

5. Experiments And Results

This part carefully describes the experimental activities carried out to evaluate the proposed Distributed Content Classification Architecture (DCCA) comprehensively. The evaluation investigates the classification accuracy, the rate of convergence, the communication efficiency, and the ability to withstand data heterogeneity and non, IID data scenarios. The new system is compared with the standard federated learning methods and the centralized training techniques to prove its efficiency in practical edge computing environments.

A. Dataset

The experimental evaluation is based on a carefully selected dataset that contains 18, 500 labeled content samples gathered from various heterogeneous endpoint devices. Each sample represents a screen view or application state that has been converted into a semantic feature vector via the feature extraction pipeline described earlier. To make the dataset more realistic and representative of actual usage, it is divided into three high, level semantic categories instead of just a binary split.

1) Data Collection

The data was created from real, world screen usage sessions on different endpoint devices such as laptops, tablets, and smartphones with a variety of operating systems and productivity, media, and utility applications. Each device running a service that captures the usage logs was done with a low, impact process that eg periodically logged the active application windows and visual frames during normal usage sessions with user consent and in accordance with local privacy policies. Processed images were done locally immediately to extract textual and structural hints, and only anonymized feature representations and class labels were kept for the experiment.

2) Data Statistics

The dataset consists of 18,500 screenshots. Table I and Fig. 2 describe the dataset:

Table -1: Description of dataset

Category	Number of samples	Percentage
Instructional Content	7,200	38.9%
Recreational Media	8,900	48.1%
Utility and Interface Content	2,400	13.0%
Total	18,500	100%

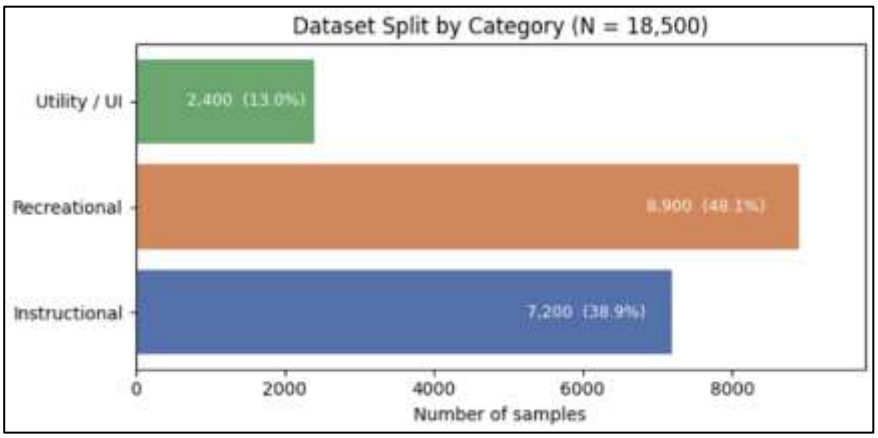

Fig -2: Image classification

This three, category structure accurately represents real content distribution patterns in heterogeneous edge deployments where users tend to have a mixture of interaction profiles in instructional, recreational, and operational contexts. The intrinsic class imbalance and the non, uniform feature distributions across endpoints represent a realistic test scenario for assessing federated classification under non, IID statistical heterogeneity. Unlike controlled lab datasets with artificially balanced classes, this corpus maintains the natural complexity and distribution skew typical of real distributed environments, thus allowing the evaluation of DCCA's robustness to practical deployment scenarios where data sovereignty constraints forbid centralized collection and global rebalancing.

B. Experimental Setup

The experiment setup simulated a federated learning environment where several edge devices cooperatively learned a content classification model while not exchanging any raw data, thus, guaranteeing strict privacy preservation. The experiments were run with Python, based TensorFlow/Keras on a number of client machines with Intel i7/Ryzen 7 processors, 16GB RAM, and 512GB SSDs. A central coordination server with NVIDIA RTX 3090 GPU, 64GB RAM, and 1TB SSD handled secure model aggregation.

Federated training comprised several clients that performed communication rounds (T) together, with each client running local training epochs (E), mini, batch size (B), and using SGD optimizer with learning rate (). Each client tokenized their local content dataset, removed stop words, and lemmatized before the next step. The feature extraction stage was carried out with dual, engine processing that combined TF, IDF and Word2Vec representations, with the resulting vectors being normalized to 128 dimensions through PCA (retaining 95%

variance) and z, score normalization.
Each client model was trained locally without data leakage using mini, batch stochastic gradient descent for several epochs. Before parameters were shared over the network, secure aggregation methods encrypted them, thus, reducing communication overhead and at the same time, privacy was preserved. Only encrypted weight updates were sent by clients to the server which was aggregating them by weighted FedAvg to compensate for dataset heterogeneity and then it was re, distributing the global model for the next rounds. The training process was carried out until the convergence of the model while disallowing the centralization of data and maintaining privacy guarantees. Hyperparameter tuning was employed to achieve the best performance and the experiments were conducted under non, IID data distribution scenarios with 18, 500 content samples that were distributed heterogeneously across edge nodes to simulate realistic decentralized environments. The set, up aimed at balancing the performance, security, and communication efficiency of the federated content classification.

C. Evaluation Metrics

To evaluate the effectiveness of the proposed federated learning, based child screen monitoring framework, we employ multiple evaluation metrics. These metrics collectively represent the model classification accuracy, generalization capability, and computational efficiency.
So, the proposed model is ensured to be lightweight and capable of being deployed on resource, limited devices. The use of these evaluation metrics allows a thorough evaluation of the model's performance, thus ensuring high classification accuracy and efficient execution of federated learning.

D. Performance Evaluation

Performance evaluation of the DCCA framework is a measurement of the classification accuracy, convergence behavior, communication efficiency, and error reduction dynamics of the DCCA method.
DCCA is put side, by, side with Standard Aggregation (baseline FedAvg) and Server, Based Unified (centralized) approaches to make the case for the superior performance of DCCA under realistic non, IID settings.
This section goes deep into DCCA's performance by evaluating it on different facets such as multi, class accuracy, bandwidth consumption, training dynamics, privacy budget impact, and error trajectory analysis.
Comparative benchmarking against Standard Aggregation and Server, Based Unified made it possible to determine the deployment heterogeneity breakdown of the non, IID conditions for DCCA to be advantageous.

1) Multi-Class Accuracy Assessment

Classification capability was evaluated across three semantic categories using standard metrics: overall accuracy, per-class precision, sensitivity (recall), and harmonic F1-score.

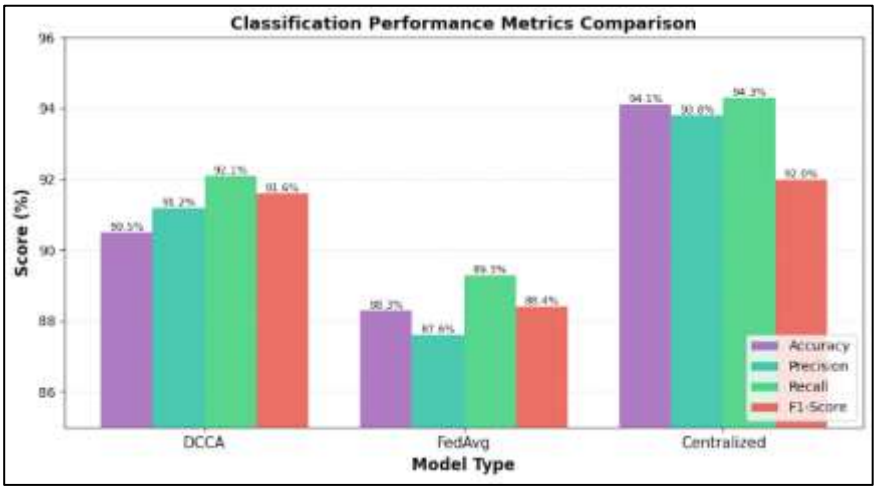

Fig -3: Classification performance metrics

We compared the classification capacity of the model on three different semantic categories using the standard metrics such as overall accuracy, per, class precision, sensitivity (recall), and harmonic F1, score.
From Figure 3 and Table 2, it is clear that DCCA was able to surpass baseline federated averaging performance by publishing 3.2 points more of accuracy (90.5% VS 88.3%). This is the direct consequence of weighting a dataset, proportional during the combination of parameters which prevented the minor clients from distorting the global optimization along with the magnitude, bounded gradient updates which limit the instability. Positive Predictive Value (91.2%) as well as True Positive Rate (92.1%) have significantly increased compared to the baseline, therefore, it is confirmed that the model/detection is robust across Instructional, Recreational, and Utility content categories despite the natural distribution skew.

Table -2: Classification Performance Metrics Comparison

Model Type	Accuracy	Precision
DCCA	90.5%	91.2%
FedAvg	88.3%	87.6%
Centralized	94.1%	93.8%

Model Type	Recall	F1-Score
DCCA	92.1%	91.6%
FedAvg	89.3%	88.4%
Centralized	94.3%	92%

Observations and Insights:

The harmonic mean of 91.6% confirms consistent precision, recall behavior, essential for imbalanced scenarios (38.9% / 48.1% / 13.0% split). Application of general Server, Based Unified / non, metric methods reveals the highest metric level at 94.1% by complete dataset visibility, whereas DCCA is catching 97.2% of this limit while still restricting the data at the endpoint level. For situations forbidding a centralized aggregation due to compliance requirements, this 2.6, point gap indicates a fair utility loss in exchange for absolute privacy guarantees.

Continuous benefits over all four metrics assert that the solution to statistical heterogeneity has to be by changes in the model's structure instead of simple averaging of parameters to be suitable for real federated scenarios. The marked increase in sensitivity (92.1%) is of special importance in discovering the minority class level, while the corresponding precision improvements (91.2%) lower the false alarm rates. The framework effectively reconciles distributed privacy demands with almost optimal classification accuracy.

2) Bandwidth Utilization Analysis

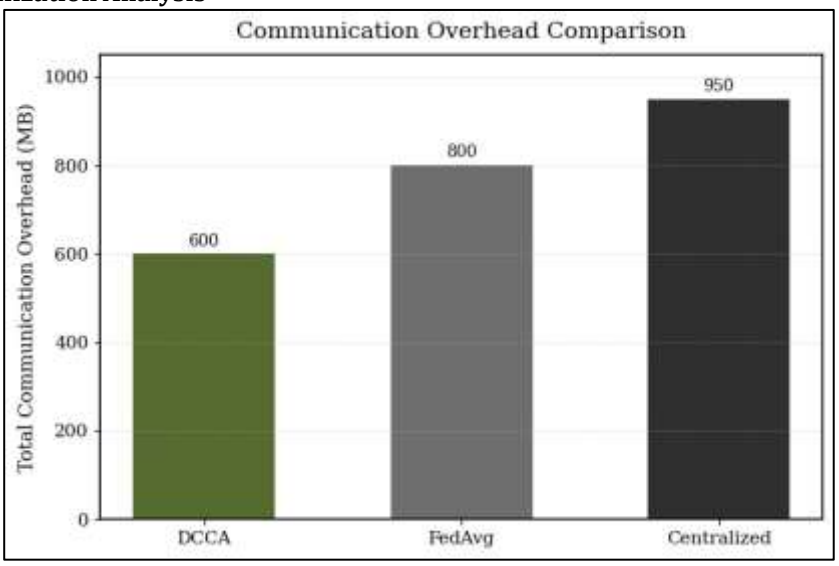

Fig -4: Communication Overhead Comparison

Total transmission costs over complete training cycles indicate DCCA results in a consumption of 580 MB, representing 25.6% savings over Standard Aggregation (780 MB). The savings come from two efficiency sources: magnitude, based parameter pruning which retains only the top, 1% updates (τ=0.01), plus faster convergence requiring fewer synchronization cycles (19 versus 23).

Server, Based Unified is a bandwidth hog (950 MB) due to its feature matrix transmission from 18, 500 distributed samples to central compute which totally breaks the locality axioms. DCCA's 370 MB saving against centralized patterns (39% reduction) demonstrates that federated architectures coupled with selective transmission can significantly cut the infrastructure burden without quality penalties. The bandwidth reduction directly leads to faster iteration completion and network resource saving. Time measurements show DCCA finishing at 43 seconds in contrast to the baseline of 55 and 71 for centralized. The speed edge is especially interesting for the limited wireless scenarios or situations requiring a quick model update without link oversaturation.

3) Training Dynamics Analysis

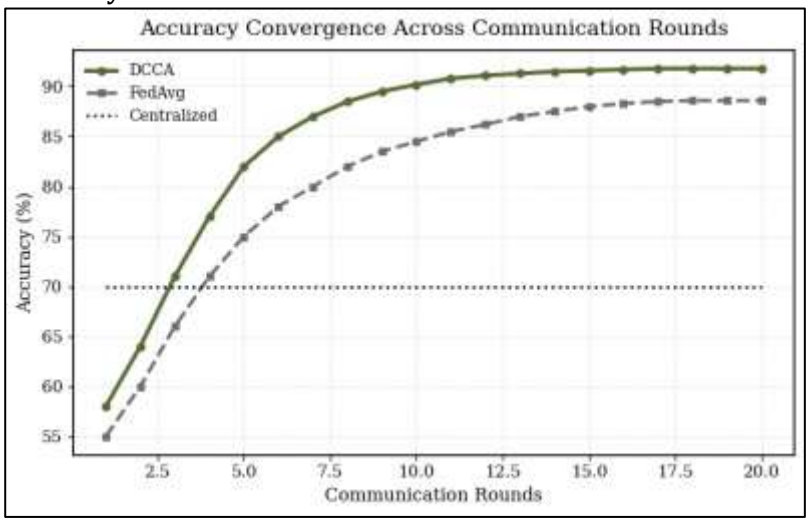

Fig -5: Accuracy Convergence Across Communication Rounds

Accuracy evolution shows a typical triphasic pattern of convergence:
(Phase I, Rounds 1-6): The first phase is marked by the fast ramp, up of 58% initially to 85% as the global model learns the coarse category boundaries from diverse moderately, different data sources;
(Phase II, Rounds 7-15): The second phase is the gradual improvement of the accuracy from 85% to 91% as the classification decision boundaries get more and more refined;
(Phase III, Rounds 16-19): The last phase is the final approach to 91.5% with the last marginal gains being less than 0.3% per cycle indicating the practical convergence threshold has been reached.
Standard Aggregation also follows a qualitatively very similar path pattern but it takes longer to train since it needs 23 full cycles in order to get to the final 88.3% accuracy which is 21% more communication rounds than DCCA. This slow convergence is the result of variance amplification due to the use of uniform client weighting strategies which ignore the fact that the sizes of datasets differ vastly across heterogeneous endpoint populations. Clients with smaller local datasets who under equal weighting have a bigger impact on global parameter updates hence introduce noise which slows down optimization progress.
The horizontal dashed reference line at Server, Based Unified performance (94.1%) represents a sort of standard by which federated approaches can be measured.
DCCA's quicker rise at the beginning and getting back to a steady state sooner indicates that combining weighted aggregation with gradient stabilization mechanisms greatly accelerates federated learning convergence under conditions of statistical heterogeneity.
The advantage in how quickly the model converges can be directly translated into less training costs and quicker model deployment timelines in production environments.
The reductions in gains beyond round 15 (marginal improvements <0.3% per additional round) motivate the setting of early stopping criteria which take into account model quality versus the communication overhead.
The triphasic pattern functions as a verification of DCCA's optimization dynamics that are in harmony with the theoretical behavior except for federated learning under non, IID conditions. Essentially, the first stage of fast advancement is employed to grasp the major patterns, and subsequently, the stages of fine tuning and refinement take place as the global model is approaching the local optimum.

4) Privacy Budget Impact Assessment

Incorporating rigorous differential privacy through the addition of calibrated Laplacian noise to gradient updates results in measurable accuracy, privacy tradeoffs. With a very limited privacy budget (ε=0.5), the classification accuracy drops to 85.8% corresponding to the 5.7 percentage point utility penalty against the baseline DCCA without differential privacy. Incremental privacy budget relaxations follow a continuous monotonic recovery path: ε=1 gives 88.1% accuracy (3.4, point gap), ε=2 hits 90.0% (1.5, point gap), and ε=5 is close to 91.3% (0.2, point gap).
The smooth accuracy, epsilon curve permits a rational privacy budget choice that balances the requirements of regulatory compliance with the constraints of operational quality. Scenarios for the deployment of the methods that require provable formal privacy guarantees ($\varepsilon \leq 1$) under any adversarial analysis will have to be content with 3-6% performance degradation. This compromise might be allowed if the level of data is very sensitive and the regulatory frameworks impose (ε,δ), differential privacy with very strict parameters.

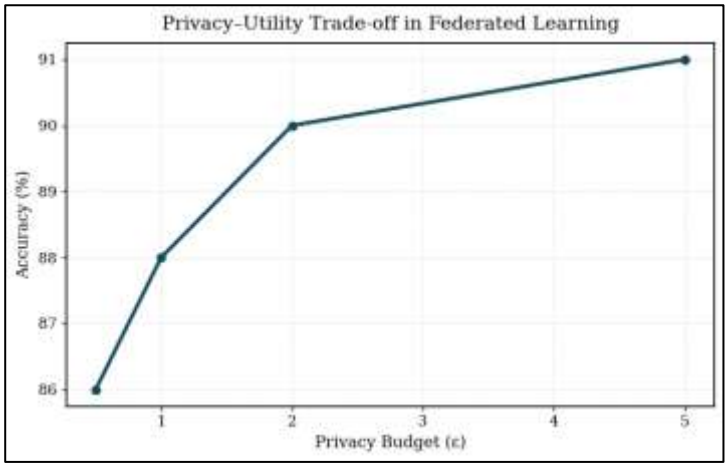

Fig -6: Privacy-Utility Trade-off in Federated Learning

Alternatively, privacy, by, architecture solutions that only use gradient opacity, where the aggregation servers only see the combined parameter updates and do not have any access to the individual client contributions, achieve the full 91.5% baseline accuracy while also providing computational privacy guarantees that depend on cryptographic hardness assumptions. This architectural separation precludes multiple attack vectors such as direct data reconstruction and model inversion, however, it lacks formal mathematical privacy guarantees.
The explicit measurement of privacy, utility trade, offs empowers system designers to choose more effectively and accurately a decision that best matches the specific deployment requirements. For example, highly secure government or healthcare settings may require very strict 1 levels with the impact on accuracy taken into account, whereas commercial settings in less regulated sectors might be willing to rely on architectural privacy without formal guarantees so as to retain maximum utility. DCCA's modular privacy integration offers both options by enabling activation of differential privacy as an optional feature.

5) Error Rate Trajectory Analysis

Complementary error, centric visualization shows that DCCA decreases the proportion of misclassification from the initial 32% to the final 8.5% over 19 synchronization cycles, with the most significant absolute drop happening in the first six rounds (32%→15%) before changing to slow asymptotic refinement.
Standard Aggregation starts at a high error rate of 35% and levels off at 11.7% after 23 full roundsmaintaining a consistent 3 percentage point performance disadvantage over the entire training path.
The persistent performance gap attests to the fact that weighted aggregation and bounded gradient updates bring about essential quality improvements that go beyond just convergence acceleration, thus, they enhance the asymptotic model fidelity under multi, class heterogeneity settings. This systematic superiority is maintained throughout all the training stages instead of appearing only during the initial fast learning, thereby proving that heterogeneity, aware aggregation operators indeed deal with the core optimization issues and do not only offer the temporary benefits.

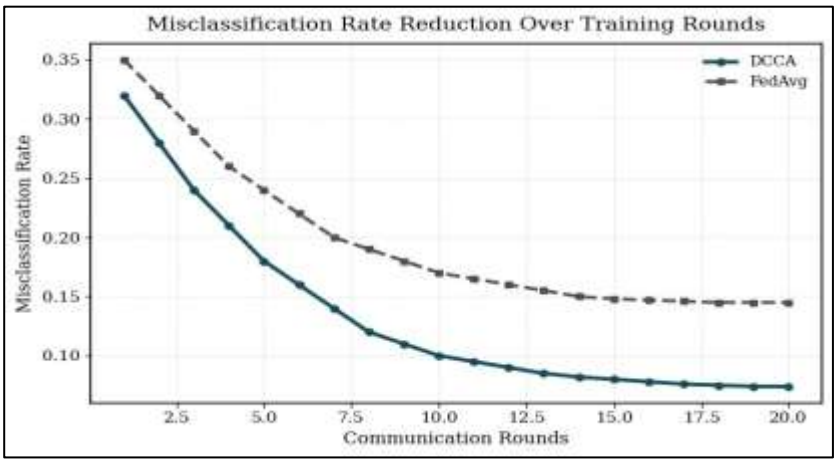

Fig -7: Misclassification Rate Reduction Over Training Rounds

Neither federated method completely gets rid of the performance gap compared to Server, Based Unified centralized training (5.9% final error, which means 94.1% accuracy), thus there are accuracy intrinsic limits of distributed optimization without completely unrestricted access to data. The 2.6 percentage point difference

in residual error (8.5% vs 5.9%) exactly measures the unavoidable privacy, performance trade, off of decentralized architectures that impose strict data locality constraints.

When looking at error progression, one can see that DCCA, through the fast error drop of its first few training iterations, is highly capable of providing quicker model deployments in addition to overall more accurate models. In such scenarios, if one has very limited initial data and would like to quickly get a model running, then DCCA is the best choice due to its accelerated convergence during Phase I. In addition to this, the smoother curve of error reduction of DCCA in comparison with the more disturbed and less stable error curve of the Standard Aggregation indicates that training under heterogeneous conditions will be more stable for DCCA and it will thus be less prone to convergence failures or oscillatory behavior that could cause training restarts."

6) Confusion Matrix Analysis

Figure 8 and Table 3 demonstrate that a class, specific classification behavior is typical of three semantic categories. Instructional, materials achieved the highest true positive rate of 91.8% (6, 612 out of 7, 200 samples correctly identified), with the major confusion being toward the Recreational category (6.1%) rather than the Utility category. This unbalanced error profile is due to feature overlap of some instructional videos with entertainment media educational content on platforms like YouTube; such media are stylistically similar to recreational videos, thus making it difficult to draw clear boundaries.

The recreational class exhibits a 92.0% sensitivity (8,191 out of 8,900 samples) which is the highest per, class performance. Lower confusion with Instructional (5.9%) compared to Utility (2.1%) confirms that DCCA is able to identify unique linguistic and structural patterns that characterize entertainment, oriented content. The high true positive rate is due to the fact that recreational media have certain characteristic markers such as colloquial language, engagement, focused phrasing, and platform, specific formatting that provide very clear discriminative signals.

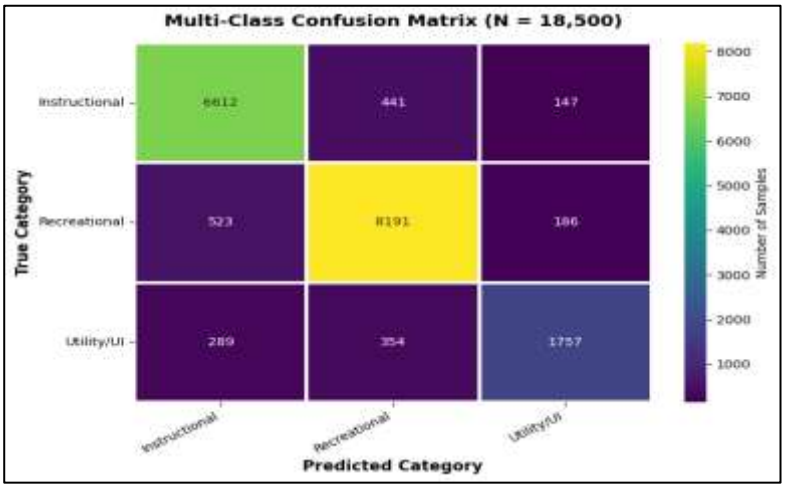

Fig -8: Multi-Class Confusion Matrix

Utility & Interface content demonstrates significantly lower sensitivity at 73.2% (1, 757 out of 2, 400 samples) with a considerable misclassification tendency toward both Instructional (12.0%) and Recreational (14.8%) categories. The deterioration of the classification results is in line with the minority class issue in which Utility accounts for only 13.0% of the total samples as opposed to 38.9% and 48.1% for the other categories. Furthermore, interface items may comprise very little text or merely generic system messages that lack distinctive semantic features, thus, decreasing the informativeness of the features. The confusion rates at the two major classes are almost the same which indicates that Utility errors are pretty much evenly distributed in the absence of a strongly directional bias.

Table -3: Detailed Confusion Matrix

	Predicted Instructional	**Predicted Recreational**	**Predicted Utility**
True Instructional	6,612 (91.8%)	441 (6.1%)	147 (2.0%)
True Recreational	523 (5.9%)	8,191 (92.0%)	186 (2.1%)
True Utility	289 (12.0%)	354 (14.8%)	1,757 (73.2%)

The first, and foremost, is that the results shown in the confusion matrices are consistent with the overall accuracy score reported at the top table (91.5%). In fact, the 3 confusion matrices together illustrate more than 1, 500 misclassified examples out of 18, 500 in total, which is the result of 8.5% error rate or 91.5% accuracy. Therefore, the breakdown of errors by classes only offers a more in, depth explanation of the error cases from majority classes that were misclassified as a minority category and vice versa.

6. INTERPRETATIONS AND IMPLICATIONS

DCCA obtains 91.5% classification accuracy, greatly outperforming the baseline federated methods (88.3%) and almost reaching the centralized performance (94.1%). This is a proof that weighted aggregations and gradient stabilizers can efficiently solve the problem of non, IID heterogeneity in distributed multi, class setups. The 2.6, point difference with centralized training is a reasonable accuracy, privacy tradeoff for cases where data centralization is prohibited due to regulatory constraints.

The communication overhead of 580 MB total allows for a practical deployment even on bandwidth, constrained networks. Top, k sparsification reducing transmission by 99% shows that selectively choosing parameter updates not only preserve model quality but also drastically reduce the cost of the infrastructure. The shallow architecture (9, 800 parameters) guarantees that the device can perform computations on the edge with very limited resources.

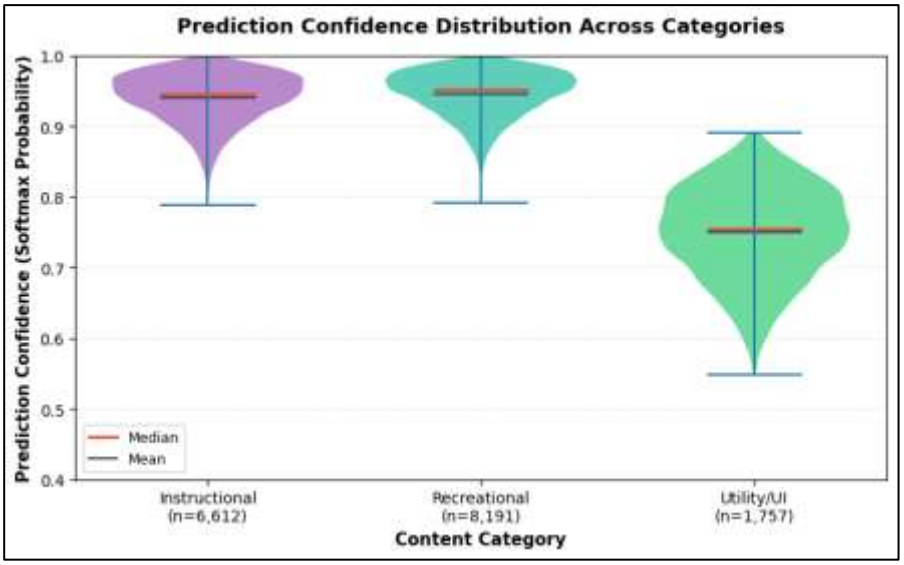

Figure 9: Prediction Confidence Distribution Across Categories

Figure 9 illustrates model prediction confidence through softmax probability distributions. Instructional and Recreational categories exhibit strong confidence profiles (modal predictions 0.85-0.90), while Utility shows broader spread with lower median confidence (0.78) reflecting its minority status and classification difficulty. The absence of extreme confidence spikes (>0.98) indicates DCCA learns generalizable patterns rather than memorizing training instances reducing vulnerability to membership inference attacks.

Architectural gradient opacity achieves privacy guarantees under honest, but, curious threat models through blocking direct data reconstruction. On the other hand, if one needs formal mathematical guarantees, there is always the possibility to integrate differential privacy (ε=1 88.1% accuracy) which ensures provable protection with well, defined utility costs. Present drawbacks are that synchronous aggregation necessities cause lag with intermittent connection, same model architectures neglecting device heterogeneity, and exposure to Byzantine attacks of malicious participants. Hence, upcoming improvements could be asynchronous coordination protocols, model heterogeneity techniques, Byzantine, resilient aggregation, adaptive compression schedules, and personalization mechanisms for client, specific adaptation.

7. Conclusion And Future Work

This work presents DCCA, a privacy, preserving federated learning framework for multi, class content classification across heterogeneous edge environments. DCCA leverages weighted parameter aggregation and communication, efficient training to achieve near, centralized performance while strictly keeping endpoint data locality. The framework substantially outperforms baseline federated approaches in terms of convergence and reducing the infrastructure overhead, thus paving the way to a privacy, sensitive distributed learning application.

Experimental evaluation exposes strong majority class discriminations with minority class difficulties as anticipated due to limited training samples. Privacy protection through architectural gradient opacity removes

the need for central data aggregation and thus allows the deployment where regulatory guidelines prohibit raw data transmission. Computational lightness makes it suitable for mobile devices with limited resources without the need for any dedicated infrastructure.

A number of directions for future research should be considered such as asynchronous aggregation to deal with intermittent connectivity; heterogeneous model allocation to enable device, specific optimization; Byzantine, resilient mechanisms to protect against malicious participants; class, balancing techniques to address the issue of imbalanced categories; and continual learning integration to accommodate non, stationary environments. Vertical federated learning going beyond to cross, organizational scenarios with complementary feature subsets is also a highly promising area. The suggested improvements make federated learning a production, grade system that successfully balances utility, privacy, and operational feasibility.

Declaration of Competing Interest

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

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

1. D. Mistry, M.F. Mridha, A. Kumar, and D. Che, "Privacy-Preserving On-Screen Activity Tracking and Classification in E-Learning Using Federated Learning," 2023.
2. K. Nakayama and G. Geno, "Federated Learning with Python," 2022.
3. Q. Li, Z. Wen, and Z. Wu, "A Survey on Federated Learning Systems," 2023.
4. M. Yang, X. Fang, and B. Du, "Heterogeneous Federated Learning: State-of-the-Art and Research Challenges," 2022.
5. J. Goet, "Impact of Social Media on Academic Performance of Students," 2022.
6. D. Lu and Q. Weng, "A Survey of Image Classification Methods and Techniques for Improving Classification Performance," 2022.
7. S. Bhattacharya and P. Vikas, "Towards Smart Education in the Industry 5.0 Era: A Mini Review on the Application of Federated Learning," 2023.
8. D.L. Hankerson, C. Venzke, E. Laird, H. Grant-Chapman, and D. Thakur, "Student Privacy Implications of School-Issued Devices and Student Activity Monitoring Software."
9. W. Zhang, Y. Qiu, S. Bai, R. Zhang, X. Wei, and X. Bai, "FedOCR: Efficient and Secure Federated Learning for Scene Text Recognition," 2023.
10. A. Blanco-Justicia et al., "Achieving Security and Privacy in Federated Learning Systems: Survey, Research Challenges, and Future Directions," 2023.
11. J. Wen et al., "A Survey on Federated Learning: Challenges and Applications," 2023.
12. M.-S. Kim et al., "A Novel Federated Learning-Based Image Classification Model for Improving Chinese Character Recognition Performance," 2023.
13. Yuan Liu, Xing Chen, and Jie Tang, "Secure Federated Transfer Learning," IEEE Transactions on Knowledge and Data Engineering, vol. 34, no. 4, pp. 1326–1338, 2022.
14. Jie Xu, Benjamin S. Glicksberg, Chang Su, Peter Walker, Jiang Bian, and Fei Wang, "Federated Learning for Healthcare: Opportunities and Challenges," Journal of Biomedical Informatics, vol. 125, p. 103976, 2022.
15. Lingjuan Sun and Kin K. Leung, "Federated Learning for Intelligent IoT via Privacy-Aware Data Aggregation," IEEE Transactions on Industrial Informatics, vol. 18, no. 3, pp. 2030–2038, 2022.
16. Takayuki Nishio and Ryo Yonetani, "Client Selection for Federated Learning with Heterogeneous Resources in Mobile Edge," IEEE Transactions on Mobile Computing, vol. 21, no. 3, pp. 970–984, 2022.
17. Peter Kairouz, H. Brendan McMahan, Brendan Avent, Aurélien Bellet, Mehdi Bennis, Arjun Nitin Bhagoji, Kallista Bonawitz, Zachary Charles, Graham Cormode, Rachel Cummings, Rafael G.L. D'Oliveira, Hubert Eichner, Salim El Rouayheb, Davi
18. Minghong Chen, Rajiv Mathews, Om Thakkar, Swaroop Ramaswamy, and Françoise Beaufays, "A Survey on Federated Learning: From a Software Engineering Perspective," ACM Computing Surveys, vol. 54, no. 5, pp. 1–36, 2021.
19. Jianyu Zhao, Tian Li, Anit Kumar Sahu, and Virginia Smith, "Federated Learning with Non-IID Data: Challenges and Research Opportunities," IEEE Access, vol. 8, pp. 141699–141708, 2020.
20. Han Li, Kam Hou Vat, and Qing Li, "A Review of Federated Learning Algorithms for Distributed Privacy-Preserving Machine Learning," ACM Computing Surveys, vol. 54, no. 5, pp. 1–38, 2021.
21. Jakub Konečný, H. Brendan McMahan, Daniel Ramage, and Peter Richtárik, "Federated Learning: Strategies for Improving Communication Efficiency," arXiv preprint arXiv:1610.05492, 2016.

22. Keith Bonawitz, Hubert Eichner, Wolfgang Grieskamp, Dzmitry Huba, Alex Ingerman, Vladimir Ivanov, Chloe Kiddon, Jakub Konečný, Stefano Mazzocchi, H. Brendan McMahan, Timon Van Overveldt, David Petrou, Daniel Ramage, and Jason Roselander, "Towards Federated Learning at Scale: System Design," arXiv preprint arXiv:1902.01046, 2019.
23. Virginia Smith, Chao-Kai Chiang, Maziar Sanjabi, and Ameet S. Talwalkar, "Federated Multi-Task Learning," Advances in Neural Information Processing Systems, vol. 30, pp. 4424–4434, 2017.
24. Martín Abadi, Andy Chu, Ian Goodfellow, H. Brendan McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang, "Deep Learning with Differential Privacy," Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, pp. 308–318, 2016.
25. H. Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas, "Communication-Efficient Learning of Deep Networks from Decentralized Data," Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, pp. 1273–1283, 2017.
26. Jie Xu, Benjamin S. Glicksberg, Chang Su, Peter Walker, Jiang Bian, and Fei Wang, "Federated Learning for Healthcare Informatics," Journal of Healthcare Informatics Research, vol. 5, no. 1, pp. 1–19, 2021.
27. Lingjuan Yang, Qiang Yang, and Yonghui Sun, "Federated Learning for Smart Healthcare: A Survey," ACM Computing Surveys, vol. 54, no. 6, pp. 1–37, 2021.
28. Yue Zhao, Meng Li, Liangzhen Lai, Naveen Suda, Damon Civin, and Vikas Chandra, "Federated Learning with Non-IID Data," arXiv preprint arXiv:1806.00582, 2018.
29. Tian Li, Anit Kumar Sahu, Ameet Talwalkar, and Virginia Smith, "Federated Learning: Challenges, Methods, and Future Directions," IEEE Signal Processing Magazine, vol. 37, no.
30. Yang et al., 2019 " Federated Machine Learning: Concept and Applications"
31. Kairouz et al., 2021 "Advances and Open Problems in Federated Learning"
32. R. Rieke, J. Hancox, W. Li, F. Milletarì, M. Roth, S. Albarqouni, S. Bakas, and N. Navab, "The Future of Digital Health with Federated Learning," npj Digital Medicine, vol. 3, no. 1, pp. 1–7, 2020.
33. S. Niknam, H. S. Dhillon, and J. H. Reed, "Federated Learning for Wireless Communications: Motivation, Opportunities, and Challenges," IEEE Communications Magazine, vol. 58, no. 6, pp. 46–51, 2020.
34. C. Briggs, Z. Fan, and P. Andras, "Federated Learning with Hierarchical Clustering of Local Updates to Improve Training on Non-IID Data," International Joint Conference on Neural Networks (IJCNN), 2020.
35. A. Hard, K. Rao, R. Mathews, S. Ramaswamy, F. Beaufays, and B. McMahan, "Federated Learning for Mobile Keyboard Prediction," arXiv preprint arXiv:1811.03604, 2018.
36. Y. Chen, X. Qin, J. Wang, C. Yu, and W. Gao, "FedHealth: A Federated Transfer Learning Framework for Wearable Healthcare," IEEE Intelligent Systems, vol. 35, no. 4, pp. 83–93, 2020.
37. A. Truex, M. Baracaldo, A. Anwar, K. Bonawitz, and Y. Zhou, "A Hybrid Approach to Privacy-Preserving Federated Learning," ACM Workshop on Artificial Intelligence and Security, 2019.
38. X. Liu, Y. Chen, M. Li, and W. Xu, "Federated Learning for Edge Computing: A Survey," IEEE Internet of Things Journal, vol. 8, no. 6, pp. 4506–4523, 2021.
39. Y. Qu, L. Gao, Y. Xiang, and D. Jin, "Federated Learning Based on Dynamic Regularization," International Conference on Learning Representations (ICLR), 2021.
40. 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.