

Article

Federated Learning-Based Analysis of Human Sentiments and Physical Activities in Natural Disasters

Muhammad Sadiq Amin and Woong-Kee Loh *

School of Computing, Gachon University, Seongnam 13120, Republic of Korea

* Correspondence: wkloh2@gachon.ac.kr

Abstract: In federated learning (FL), in addition to the training and speculating capacities of the global and local models, an appropriately annotated dataset is equally crucial. These datasets rely on annotation procedures that are error prone and laborious, which require personal inspection for training the overall dataset. In this study, we evaluate the effect of unlabeled data supplied by every participating node in active learning (AL) on the FL. We propose an AL-empowered FL paradigm that combines two application scenarios and assesses different AL techniques. We demonstrate the efficacy of AL by attaining equivalent performance in both centralized and FL with well-annotated data, utilizing limited data images with reduced human assistance during the annotation of the training sets. We establish that the proposed method is independent of the datasets and applications by assessing it using two distinct datasets and applications, human sentiments and human physical activities during natural disasters. We achieved viable results on both application domains that were relatively comparable to the optimal case, in which every data image was manually annotated and assessed (criterion 1). Consequently, a significant improvement of 5.5–6.7% was achieved using the active learning approaches on the training sets of the two datasets, which contained irrelevant images.

Keywords: federated learning; active learning; machine learning; human visual sentiment analysis; human activity analysis



Citation: Amin, M.S.; Loh, W.-K. Federated Learning-Based Analysis of Human Sentiments and Physical Activities in Natural Disasters. *Appl. Sci.* **2023**, *13*, 2925. <https://doi.org/10.3390/app13052925>

Academic Editors: Moongu Jeon and Jeonghwan Gwak

Received: 9 January 2023

Revised: 18 February 2023

Accepted: 21 February 2023

Published: 24 February 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Federated learning (FL) is a cutting-edge artificial intelligence approach that performs concerted training of local machine learning (ML) models across decentralized nodes, while preserving the data privacy of each edge device. In recent years, FL has been intensively researched and implemented in a variety of domains in which data privacy is a key concern. In addition to several other limitations, the feasibility of FL in an application depends on the provision of high-quality labeled datasets at each client node participating in efficient training [1]. In particular, supervised ML is hindered by the need for human annotator involvement in image data gathering and annotation to prepare the training data. In general, this procedure requires increased manpower or crowdsourcing to analyze and label the dataset for model training. First, each sample must be meticulously examined, which is a laborious and tedious task. Second, this approach does not ensure variety in the quality of the images with sufficient detail patterns, which is vital for the efficiency of ML models. Both these challenges are significant obstacles that must be overcome for the approach to be successful. A potential solution to these constraints is active learning (AL), which is a learning paradigm that allows learning techniques to be probed concurrently as an information source when choosing and identifying future training datasets. Deep models require a large amount of training data to optimize a large number of parameters, which is important for the model to learn how to retrieve high-end features; hence, deep models are data voracious. In such a scenario, AL enables a learning system to proactively explore source data while simultaneously tagging incoming training data, which may provide solutions to the highlighted constraints. It simplifies the annotation process by

labeling data from a large unlabeled data pool in an automated manner using ML models that are trained on a significantly reduced hand-labeled dataset.

The existing FL literature suggests that each client node is exposed to predefined manually annotated (supervised learning) training image sets. However, in some circumstances, for instance during semi-supervised and unsupervised learning, each client node may possess a vast amount of unlabeled data that can be utilized to train local models, thereby enhancing the efficiency of the global aggregation model. In this article, we present an innovative paradigm for FL that is empowered by AL to interactively aggregate a global model, while considering privacy protection in a multi-stake context using unlabeled data accessible at each client node. To achieve this, we used and assessed a range of AL techniques with diverse sampling and disagreement approaches. Specifically, two distinct sampling and disagreement procedures were investigated in detail using two pool-based approaches. First, the labels of those instances for which the current prediction is maximally uncertain are sequentially queried through uncertainty sampling. This approach mitigates several downsides, including the fact that uncertainty sampling tends to be skewed towards the legitimate learner and might overlook key instances that are hidden from the estimator's perspective. Second, this issue is addressed using query by committee (QBC), which involves maintaining many hypotheses consecutively and selecting queries among them when discrepancies appear.

In the proposed approach, we first keep the AL procedure offline to prevent complexity in terms of communication and learner biases at each node, with communications with the server initiated after the dataset at each node is labeled. At the onset, offline AL limits the number of communication cycles during FL and restricts the selection of data in every cycle. Furthermore, because the global aggregator model acts as a learner in this scenario to acquire and label training datasets during FL, for instance at every communication cycle, it potentially impedes the convergence of the global aggregator model owing to the data sampling bias. In this scenario, there are two primary obstacles. Initially, if the sample size to be selected for every cycle is kept extremely large, the global aggregator model efficiency will become more variable. Secondly, maintaining a smaller sample size may expand the number of communication cycles. In addition to these problems, an additional challenge exists for FL based on online active learning, it needs a wider array of manually labeled seeding training data that is distributed to an entire row of client nodes during the training of the models.

Two fascinating applications are used to assess the proposed approach: (1) the analysis of human visual sentiments in the social media imagery of natural disasters, and (2) the physical activity connected with the natural catastrophe response. In addition to the novel methodology, this study also investigates a distinct perspective of the applications when compared with the current studies on the subject, comprising our prior contributions, as discussed in detail in Section 2. This study provides a solid foundation for future research on this topic. The following are the most significant contributions of this study to the ongoing research in this domain:

- We investigate the viability of autonomous annotation of training data for FL using a novel methodology for generating a global model from locally available unlabeled data (on each node device).
- We demonstrate that AL can achieve equivalent outcomes in federated and centralized learning without requiring human intervention to label the data.
- We demonstrate that the efficacy of FL may be adversely influenced by a lower amount of data for the training of local models. Nevertheless, AL may be effective in such cases to acquire relevant data without any manual annotation.
- In two distinctive contexts, we assess the performance of two AL techniques based on a pool with five alternative disagreement sampling approaches under centralized and FL, in which the training of a model is performed by uploading all participating nodes' data to a cloud aggregator.

The rest of the paper is structured as follows: Section 2 covers published studies in four distinct areas, namely, FL, AL, human sentiments, and associated physical activities in disasters based on ML, notably deep learning. Section 3 covers the suggested approach for active learning-based FL using the human sentiment accompanying an associated physical activity analyzer in disaster-related visual imagery. It examines the experimental outcomes, including assessment trade-offs. Section 5 provides valuable lessons from the experimentation. Section 6 summarizes the aims and addresses future research prospects in this research area.

2. Related Work

This section provides a comprehensive overview of FL, AL, and two unique applications for the proposed framework assessment: (1) analysis of human sentiments in disastrous circumstances, and (2) human physical activities in disastrous situations.

2.1. Federated Learning (FL)

The recent research on FL primarily emphasized the challenge of optimizing a global model with non-IID (non-independent and non-identically distributed), imbalanced, and widely dispersed data, and ensuring privacy and communication performance [1]. In an FL context, McMahan et al. [2] proposed FedAvg, a global optimization approach that efficiently combines the parameters of local models to cope with unbalanced and non-IID data forms. To reduce the number of communication rounds, the framework selects a subset of clients rather than all participants in each iteration. One of the key shortcomings of FedAvg is its inability to handle varied data. Li et al. [3] presented FedProx, a revamped version of FedAvg that guarantees convergence in heterogeneous networks to handle diverse data sources in the FL context. To overcome this problem, one proximate factor was subsequently introduced into the optimization problems of the model to account for the heterogeneity caused by incomplete data. MOCHA was presented by Smith et al. [4] and has shown promising results in coping with statistical hurdles associated with FL by using multitasking, instead of training a single global model, as is the case with existing approaches, multiple global models were trained for every node. A major part of the literature is concerned with the integrity of model updates. There are two types of privacy in FL: local and global [5]. The first aims to safeguard the local model parameters, whereas the second preserves the parameter privacy of global models.

Bonawitz et al. [6] proposed a safe multiparty computation framework to protect the model on node upgrades from the aggregator model, whereas the aggregator upgrades with the final parameter to node models. Konečný et al. [7] introduced a differential privacy approach to FL to safeguard model updates.

For training, in contrast to conventional ML methods, FL does not require the transmission of data from client nodes to a centralized server. Local learning models were implemented on end devices, and their parameters were transferred to the edge/cloud server for global model aggregation through edge communication. After aggregation, the global model parameters were then transferred to the end devices [8]. The implementation of FL is fraught with difficulties, including the need for a significant amount of computational power and backup energy to compute a local training model. Training a local model to enhance learning accuracy requires a large volume of data [9]. However, for computationally restricted devices with limited datasets, it is challenging to develop local models with enhanced accuracy. We suggest that this is an interesting research subject that should be pursued to enhance global model efficiency.

2.2. Active Learning (AL)

The research on imagery, text, video, and multimedia retrieval has made extensive use of AL [10]. To obtain the desired classification results, a learning system may work in tandem with the user through AL. Using AL, a chunk of unlabeled data can be randomly selected for further classification by the algorithm. This technique is premised on the notion

that if an ML algorithm is empowered to choose the data for its learning, it may achieve a higher degree of accuracy, while using fewer labels for training. Sun et al. [11] used AL to perform context-aware image labeling by utilizing related metadata that contains four distinctive features: spatial information, camera tags, client labels, and timestamping. Clustering uses these four distinct features to categorize images into separate labeled classes. Liu et al. [12] proposed a study in which active AL reidentifies people using an early AL paradigm that labels multiple images as opposed to a single image. As its name suggests, this approach is used in the first phases of an analysis in which there are neither reference nor pre-labeled images available for annotation. In another study [13], an AL method for content-based visual extraction was presented. It has a rating model that uses support vector machines and a similarity metric that compares images queried from a dataset. By contrast, a study [14] described an AL method based on multiple criteria for dynamically labeling images using a convolutional neural network (CNN).

In other demanding applications, such as hyperspectral image classification, where reduced images are available for training deep learning models, the implications of AL are effective. Cao et al. [15], for example, suggested a CNN empowered by AL to be trained on reduced labeled images, which was utilized to annotate related images from an unannotated pool of image datasets. For classification images captured by spectral camera sensors, AL methods are used to integrate spatial and spectral information in a fused manner [16]. This research [17] employed a fusion algorithmic framework based on AL that classifies a hyperspectral image dataset by integrating spectral and temporal data. We presented the study and the recommended proposal for the two cases described in Section 3, based on the demonstrated effectiveness of AL techniques in such pertinent and complex circumstances.

2.3. Human Sentiment and Associated Activity in Natural Disasters

In sentiment analysis, ML techniques involving text analytics and biometrics are used to comprehensively recognize, isolate, assess, and analyze emotional states. Studies on inferring sentiments from image features are limited [18]. In addition, because this is a novel and intricate endeavor, the public dataset is reduced, making it challenging to establish firm state-of-the-art that can serve as a benchmark. A study proposed by Zhu et al. [19] identified low- to high-level features from the background of objects in pixels to reduce the effectiveness gap. The recognition performance was improved owing to the retrieved features. Priya et al. [20] suggested a model that reduces the effectiveness gap between the retrieval of objects with high and low background feature levels. These extracted features enhance the classification accuracy. The research described in [21] extracted features to classify the emotional reactions elicited by an image using art theory and psychological principles. The retrieved features were organized into categories according to their content, texture, and color before being fed into a naive Bayes classifier for classification. However, the retrieved features failed to discern the intricate association between human emotions and the contents of the images, even though the research produced some encouraging findings at one point. Figure 1 shows the taxonomy of human sentiment analysis using ML and deep learning algorithms.

Recent research has shown the criteria necessary to extract adjective–noun pairs (ANPs), which may be used to infer the emotional mood conveyed by a picture [23]. Examples of ANPs are “sad face” and “happy wife.” To make a meaningful contribution to the body of existing research, they assembled a dataset that included 3000 ANPs. In addition, they provided several baseline models that were utilized to assess APN-based approaches [24]. These models have been utilized extensively. Zaid et al. [25] proposed a method for predicting sentiment avatars from images. This method was a part of their study. They trained SimleyNet [26] using a dataset that had been obtained primarily from social media. This dataset includes four million images and emoji pairings. Several experts have proposed that sentiment classification should also consider data from visual and textual

sources. To analyze sentiments, for instance, the authors [27] presented attention-based modality-gated networks, which made use of imagery that went along with the text.

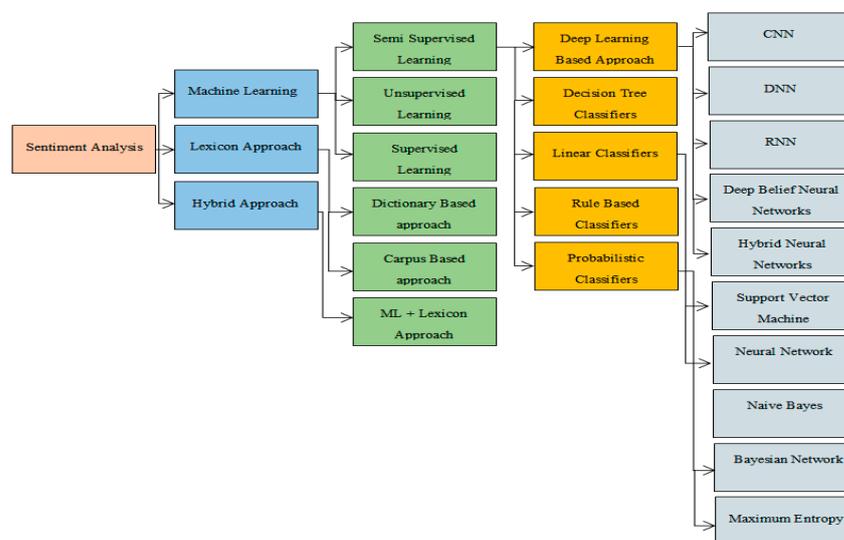


Figure 1. Taxonomic classification of sentiment analysis [22].

The most prevalent high level indications for human activity recognition in images are visible organs that can be easily moved, such as hands, legs, and neck, and the features of a background. Most published research on human activity detection is based on supervised learning [28]. In the context of gesture recognition, deep models require substantial amounts of data to be used during training, and as a solution to this issue, the transfer learning strategy has been extensively researched [29].

Because many satisfactory results have been produced, it is generally accepted that it is highly expensive to annotate all human actions. Human annotators and deep learning experts have made varying endeavors, and the possibility of error remains significant [1,30,31]. Nonetheless, providing accurate and pertinent information is imperative for the detection of human physical actions.

Crowdsourcing has been conducted to ascertain and annotate human behavioral patterns around natural calamities and associated photos on social networks. Owing to the lack of available datasets and the profusion of unlabeled images posted on social media sites, this is a potential approach to evaluate the proposed research. On the one hand, the suggested framework enables us to circumvent problems associated with the lack of training sample availability. On the other hand, it will allow blended learning in a plurality scenario without data sharing, resulting in an increase in data privacy.

3. Methodology

The proposed method for our AL-based FL approach is illustrated in Figure 2, which comprises three major modules: extracting features from images, the AL block, and the final module based on the FL block. The procedure commences with the extraction of features from the input images using our suggested CNN that is trained on the pedestrian dataset [32]. We suggest a framework [22] that focuses on human emotions in a socially significant sphere, specifically social media crisis-based visual analysis, along with an analysis of human actions. A similar approach was used in this study [33] to understand the impact of AL with FL on waste management. However, our approach utilizes CNN-based classification. We designed the architecture of this CNN that was trained on the [32] dataset with the sole purpose of classifying human activities and the associated visual sentiments, which is consistent with our previous study [22]. The motivation for selecting this CNN is based on [22]. Following this, a reduced labeled dataset, referred to as a seed, is used for the training of a classifier.

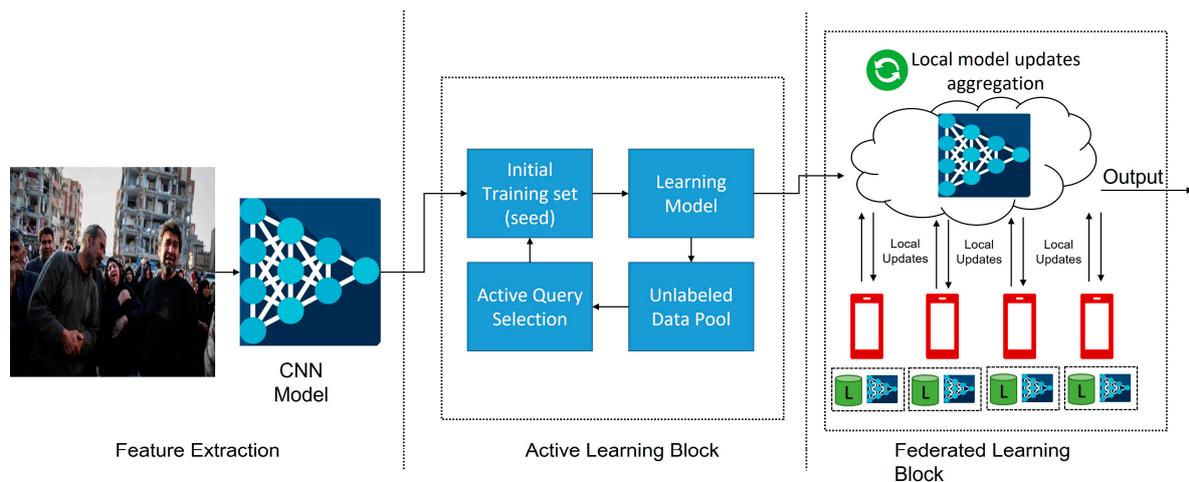


Figure 2. Infographic illustrating the proposed AL-based FL architecture.

This learner (classifier) is then employed for the selection and annotation of the image-wise range of the unlabeled image pool, using a diverse range of sampling and disagreement techniques in an iterative manner. The AL procedure proceeds until a suitable image for learning is obtained from the unlabeled image collection. The gathered training data are subsequently utilized in the training of local models at participating nodes/devices, which are ultimately integrated to establish a global aggregator.

3.1. Extraction of Features

We utilized a conventional approach for feature extraction because the primary objective of this work was to determine where FL may leverage unlabeled data delivered to each collaborating client node through AL. To accomplish this, for deep feature extraction, a CNN was utilized to identify over 1 million images and 1000 individual pedestrian crossings in a huge individual re-identification dataset. Table 1 lists the CNN architecture used to monitor and count people. We constructed a large residual network with six residual blocks and two convolutional layers. In the dense layer, the global dimensionality feature map 128 was retrieved.

Table 1. CNN architecture of the deep feature extractor convolutional neural network.

Layer	Patch Size	Stride	Output
Conv	3×3	1	$32 \times 128 \times 64$
Conv	3×3	1	$32 \times 128 \times 64$
Max Pool	3×3	2	$32 \times 64 \times 32$
First residual block	3×3	1	$32 \times 64 \times 32$
Second residual block	3×3	1	$32 \times 64 \times 32$
Third residual block	3×3	2	$64 \times 32 \times 16$
Fourth residual block	3×3	1	$64 \times 32 \times 16$
Fifth residual block	3×3	2	$128 \times 16 \times 8$
Sixth residual block	3×3	1	$128 \times 16 \times 8$
Dense	-	-	128
Batch and l2 normalization	-	-	18

The last batch, followed by l2 normalizations, maps the features on the unit hypersphere, such that they are in line with the cosine-appearing parameter. Our selection of the CNN to extract the features is influenced by the current research on two domains [2]. In particular, the feature extraction components of the AL and FL sections are self-reliant. Thus, the model used to extract the features does not affect the general analysis and intuition of AL empowered by FL.

3.2. Retrieval of the Training Dataset through AL

The primary objective of the AL stage is to obtain and label training data from a pool of unlabeled data present on local nodes. This step is performed by AL approaches based on the pool in which a model, such as “ α ”, chooses and labels the data to train from a collection of unlabeled data, such as $f = \{a_b\}_{b=1}^n$. To perform this, the model α is originally constructed using a smaller, manually labeled dataset called “seed.” We primarily applied (a) uncertainty sampling and (b) QBC, which are both pool-based approaches. The uncertainty sampling approach enables the model to evaluate relevant instances that might be obtained from a collection fuzzy environment (i.e., the degree to which the learner is uncertain in categorizing a label to the data). By contrast, the committee inquiry depends on many hypotheses/learners to select an instance, and learning models conclude the prediction based on disagreements. Various selection and disagreement methods are used to assess both approaches. The primary purpose of evaluating the approaches under various sampling and disagreement approaches is to provide a comprehensive evaluation of the existing approaches, which is anticipated to be a benchmark for future works on the topic. QBC and uncertainty sampling are assessed using the following disagreement approaches, each of which are outlined below in detail.

- Margin sampling:

As indicated in Equation (1), this method aims to pick an instance with the least variance between the probability of the two classes. Where d represents a sample of the data, and x'_1 and x'_2 are the two labels.

$$Zx(d) = p_{\alpha}\left(\frac{x'_1}{d}\right) - p_{\alpha}\left(\frac{x'_2}{d}\right) \quad (1)$$

- Least confidence:

This approach selects instances with minimal confidence scores assigned by a learning model. Equation (2) calculates the least confidence score, where the selected instances are denoted by d .

$$Z_{lc}(d) = \arg_d \max - p_{\alpha}\left(\frac{x'}{d}\right) \quad (2)$$

- Entropy:

The vote of entropy is a QBC extension that tallies the number of votes on the label from the committee. This is represented by Equation (3). Here, x_i denotes all potential labels, D reflects the committee classifiers, and $v(x_i)$ indicates the number of learners/classifiers that predict the label x_i .

$$NC_{VE}(d) = \arg_d \max - \sum_i v(x_i)/D \log v(x_i)/D \quad (3)$$

- Entropy sampling:

This approach selects instances that have maximum entropy. Equation (4) provides the detailed calculation, where $P(y|x)$ depicts the posterior probability, y represents the uncertainty metric, and Z_{es} provides the output.

$$Z_{es}x = - \sum_{y \in Y} p_{\alpha}\left(\frac{y}{x}\right) \log 2 p_{\alpha}\left(\frac{y}{x}\right) \quad (4)$$

- Maximum disagreement:

This approach quantifies the degree of disagreement between each classifier and the agreed probability and selects the sample with the greatest degree of variance for each learner.

As the first step in the AL module, a seed is fed into a model for training, which contains a tiny set of instances that have been manually labeled. The model predicts the

labels of other instances in an unlabeled image collection, where the labels are added to the seed based on a criterion described in the sampling and disagreement strategy that underpins the AL component. In particular, this is a repetitive process in which pertinent instances from the collection are retrieved. This procedure continues until all the instances from the collection are retrieved. The maximum number of repetitions is an important criterion that must be specified because the relevance of the sample that was selected will begin to decrease after a given number of iterations; thus, at some point, the AL will be forced to include samples that are irrelevant in the training set. To achieve this objective, several methods can be adopted to set the number of repetitions. Ideally, the procedure may be terminated when the learning accuracy achieves a sustainable level. Algorithm 1 depicts the termination condition that is based on the maximum repetitions.

Algorithm 1. Retrieval of training data and labeling using AL for each participant node model.

Expectation: Images as the input, classification algorithm, and sampling or disagreement mechanisms for query selection

Requirement: Image labels from the data pool

Initialization: Data distribution for feeding as the seed and unlabeled image data pool

for $j = 0:n$ **do**

Step 1: Training of the model fed with the seed

Step 2: Sample label prediction and seed selection/querying for samples in the unlabeled pool

end for

3.3. Aggregator in the FL Context

The FL architecture is the third module of our architecture, which is influenced by the federated averaging concept [34]. The stochastic gradient descent of every participating node model was aggregated to establish an aggregator global model. The fundamental design of the FL module is shown in Figure 3, which shows that the parameter α_t of the aggregator model is distributed to all collaborating nodes by the main aggregation server [35]. The procedure begins with the servers initializing global parameters that are subsequently shared among k clients. The clients then upgrade the settings using locally stored data and transmit the information back to the server, which repeats the procedure. The t represents the initial global model parameters, while K , n , n_k , t_k , and $(tC1)$ depict the cumulative number of clients, the size of the entire data, the data size of the k -th client, the parameters of the local model trained by the k -th client, and the updated global model parameters at time t , $C1$, respectively. Each node model then starts training native models based on the parameters that they retrieve from the global aggregator model. The fundamental design of the FL module is shown in Figure 3, which shows that the parameter α_t of the aggregator model is distributed to all the collaborating nodes by the main aggregation server. Each node model then starts training the native models based on the parameters that they retrieve from the global aggregator model. The parameters of local collaborating clients or nodes (e.g., α_t^k) are shared after the completion of local training with an aggregating server for updated parameters, for example, α_{t+1} , by iterating the procedure using the parameter distribution of the updated model with the nodes. This procedure is repeated for a specific transmission cycle.

In this implementation, seven client nodes engage in FL. The learning approach is a recurrent neural network [36] with long short-term memory (LSTM) [37]. Importantly, the LSTM was trained using the feature representations retrieved using the CNN elaborated in Table 1, as previously explained. We are encouraged to select LSTM because training a CNN after every session to update the network as FL, requires substantial computational hardware at the nodes. As mentioned in [38], the combination of the LSTM and CNN assists in integrating the performance for improved classification. To design the global model using the aggregation of the parameters of the local models, we used optimization algorithms, specifically FedAvg. The main FedAvg algorithm is presented in Algorithm 2, which contains two main modules. The first module contains operations related to the

server, whereas the second module contains operations specifically related to the client nodes. Here, α_t is the parameter related to the global aggregator, and the number of node devices is denoted as α^k , k depicts the batch size, B represents the number of iterations, the learning rate is denoted as E , the volume of the dataset is denoted as η , and the size of the dataset at the node device is represented as n_k . These parameters are associated with each collaborating device (i.e., α_t^k).

In addition, the fine details of the FL aggregator, such as the communication cycles with nodes, total number of nodes, and total number of active connected devices per iteration, are presented in Section 4.

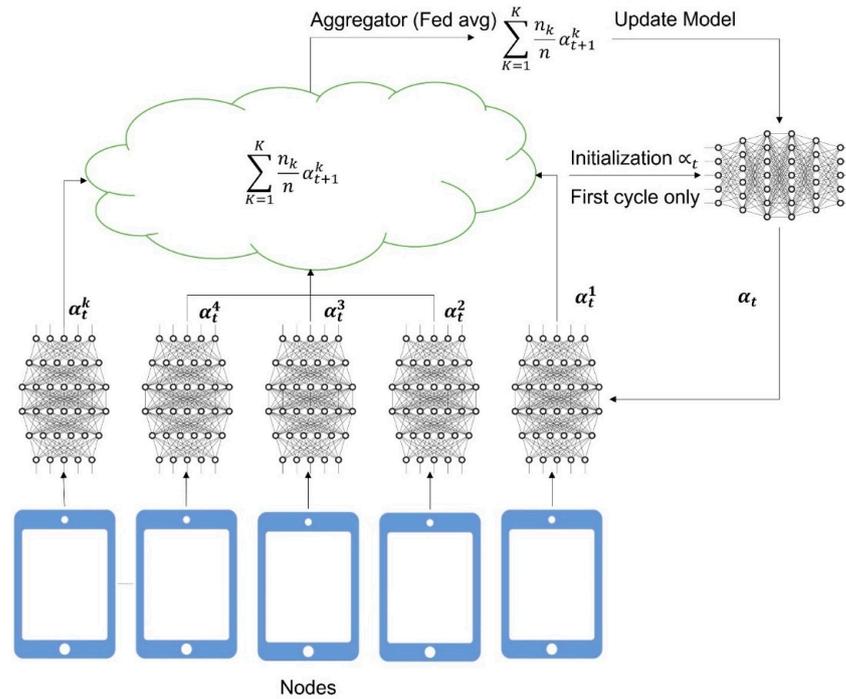


Figure 3. Diagrammatic representation of the FL architecture [35].

Algorithm 2. Federated aggregative model inspired by the FedAvg method for K client nodes.

Requirements: B, K, T, n_k, η and n

Ensure: α^t parameters of the global model

Tasks performed at the federated aggregator end:

for every communicative cycle $t = 1, 2, 3, \dots$ do

Step 1: Selection of a group of connected nodes $m = C \times K$, where $C \in (0, 1)$

Step 2: Downloading of α_k into every connected node k

for node $k \in m$ do

Step 3: Synchronize each node device

Step 4: Calculate $\alpha_t = \sum_{k=1}^m \frac{n_k}{n} \alpha_k$

end for

end for

Client devices (assume client at K) operations:

$\alpha^k = \alpha_t$

for every iteration 1 to T do

for $b \in B$ do

$$\alpha^k = \alpha^k - \eta \nabla L_k(\alpha_k, b)$$

end for

end for

return α

4. Experimental Results

4.1. Dataset

The first step begins with the collection of visuals from social media sites, such as Facebook, Google Images, and Flickr. During crawling for imagery from the aforementioned websites, the copyright element was considered, and only visuals authorized for distribution were selected. In addition, visuals were retrieved for tags, such as storms, seismic events, hurricanes, and tidal waves. These tags were made specific for places, such as “earthquake in Nepal,” “monsoons in Brazil,” and “natural disasters in Japan.”

The identification of labels for emotions and related behaviors is among the most crucial concerns of this research. The majority of the current works focus on distinct emotions, such as “negative,” “positive,” and “neutral,” with no human-related behaviors [39]. Nevertheless, we want to focus on emotions that are highly pertinent to disaster-related details, for instance, labels like “grief,” “excitement,” and “rage” in catastrophic conditions. Moreover, based on recent research in human psychology [40], we argue that persons who are surrounded by a catastrophe seem to be more likely to express two important sets of human emotions. This investigation revealed various human emotions [29]. The first collection of emotions comprises common human phrases, such as “negative,” “positive,” and “neutral.” The second category has the terms “pleased,” “anxiety,” “normal,” and “worried.”

The final collection is the expanded version of human expressions, which contained additional specific emotions, such as “anguish,” “revulsion,” “delight,” “wonder,” “excitement,” “sadness,” “pain,” “weeping,” “scar,” “antsy,” and “relief.” These emotions are related to the terms “seated,” “stand,” “ran,” “lying,” and “jogging.” Table 2 provides a thorough classification of potential human emotions and related human actions under catastrophic situations, and Figure 4 illustrates the selection of images extracted from the dataset. The dataset is made publicly available for academic research [41].

Table 2. Comprehensive spectrum of human emotions and corresponding actions utilized during the stage of crowdsourcing.

Set	Sentiments	Physical Activities
Set 1	Neutral, negative, and positive	
Set 2	Cheerful, sorrowful, calm, and frightened	Run, sit, lay, and walk
Set 3	Rage, contempt, delight, astonishment, excitement, grief, pain, tears, panic, anxiety, and relief	

Using individual viewpoints and thoughts on disasters with accompanying external perceptions, the crowdsourcing initiative aims to generate an objective for envisaged human sentiment analysis. During the crowdsourcing research phase, we provided preselected images to the study participants for annotation through Fiverr [42]. Individuals were labeled with 500 images throughout the procedure. We disseminated images with a questionnaire survey, including a disaster-related image, to participants in order for them to label human emotion and associated behavior. In the initial query, participants were asked to evaluate images on a scale from 1 to 10, where 10 represents ‘positive’ emotion, 5 represents ‘neutral’ emotion, and 1 represents ‘negative’. The purpose of this inquiry was to ascertain the respondents’ general impression of the image. The follow-up question is somewhat more precise, such as ‘sad’, ‘happy’, ‘angry’, and ‘calm’, and can retrieve the exact emotions represented by an image to respondents. In the third question, respondents were asked to rate the imagery on a scale from 1 to 7 and to describe the emotion expressed by each image.

In addition, the participants were requested to describe their sentiments about the presented images and explicitly tag images with a specific sentiment tag if that tag was not included in the list of available tags. The fourth question seeks to identify image characteristics that elicit human emotions and behavior at the threshold of the scenario or

underlying context. In the fifth query, respondents were asked to provide their opinion about a human activity, such as ‘seated,’ ‘stand,’ ‘running,’ ‘walking,’ and ‘jogging.’ Table 3 shows a concise distribution of images in the dataset.



Figure 4. Sample images for human sentiment and related physical activities in the natural disaster dataset [22].

Table 3. Detailed statistics of human sentiments and related physical activities dataset. The dimension is kept at 224×224 for every image during training.

		Number of Images					
Sentiments	Positive	518	Happy 413	Excited 105	Relief 200	Surprised 180	
	Negative	2002	Fear 608	Anger 92	Sad 1123	Disgust 203	
	Natural	480					
		Number of Images					
Activities	Standing	780	Walking 782	Jogging 32	Running 708	Sitting 780	Laying 17

At least six people were chosen to examine the images to confirm the uniformity of the labels. During the analysis, 10,000 unique responses from 2300 unique individuals were gathered. The participants represented a wide range of ages, genders, and 25 different regions. The latency for a person was 200 s, which allowed us to weed out sloppy and incorrect responses. Cohen’s kappa (k) is used for inter-rater reliability, which was calculated as 0.65. Before the final analysis, two trial tests were conducted to refine the test, fix discrepancies, and increase consistency and clarity. The datasets were split into test and training sets, with 5500 for training instances and 1695 for test instances, according to the standard protocol of breaking down the dataset by 70% and 30% for the training and test sets, respectively.

4.2. Experimental Setup

Multiple experiments were performed to demonstrate the efficacy of the proposed AL-based FL framework. On the one hand, we tested and evaluated the results of the AL

techniques in the FL environment against two baselines: Baseline 1, a carefully annotated training set; and Baseline 2, a sparsely annotated set comprising impurities. Because the objective of this study is to assess the advantages of AL in an FL environment, comparing the two baselines rather than the service-oriented architecture (SoA) in both domains is highly practical. The first standard depicts the best-case situation in which manually labeled training data exist, whereas the second criterion depicts the worst-case situations in which a model is trained on a dataset that contains a significant amount of inconsequential data. To do this, the dataset for the human sentiment criterion was synthesized by including up to 35–40% unrelated images in the collection of unlabeled data. The human physical activities in the natural catastrophe analysis application, on the contrary, provide a more realistic situation in which the second baseline is trained on a collection of social media images with the accompanying tags/queries, without human inspection and elimination of unnecessary image data. However, for the hand-labeled baseline, each picture was carefully evaluated and labeled using crowdsourcing.

In addition, one of our goals was to demonstrate how the effectiveness of the AL approaches varies depending on whether they are used in FL environments.

Furthermore, we investigated how clients with small sample sizes impact the performance of a global model by performing the analysis. Throughout the investigations, the experimental setup for AL and FL was maintained in their original state. In the next subsections, we describe in detail the experimental procedures performed using AL and FL.

We have included a summary of the parameter values utilized throughout the experimental procedure in Table 4.

Table 4. Important parameters utilized throughout the experiment.

Parameters	Values
Max AL iteration in human sentiments and related physical activity case	2500
Total client nodes	6
Maximum communication cycles	52
No. of nodes linked per cycle	6
No. of training epochs	50
Batch size	32
LSTM layers	2
No. of neurons in LSTM layer 1	100
No. of neurons in LSTM layer 2	20

The seeding images were determined based on the number of manually annotated training images available for an application. There are many variables to consider while training a learner, such as the amount and quality of the data in a seed. AL's fundamental premise is that there is a trade-off between the amount of effort needed to label the data and its effectiveness, which can be seen in the first training dataset. The ability to obtain better outcomes with fewer seeds is critical for the success of AL techniques.

We started experimentation to feed as a seed with 180 instances, which included 18 images from every training image category of human sentiments in a natural disaster. Meanwhile, we used 140 instances that contained 18 images from human physical activities in the disaster dataset. The image instances were then enlarged to pertinent instances by selection using the query strategy in every repetition. Manual annotation of the test and seed sets is essential for vigilance.

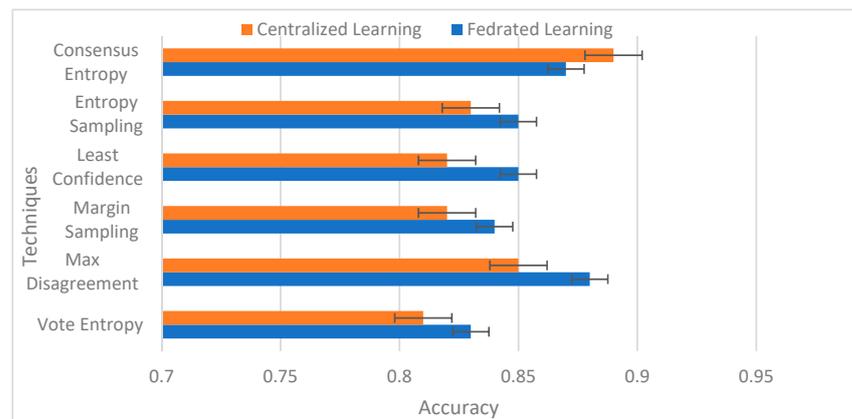
In a manner similar to that of the active learning component, a predetermined experimental setup was used throughout the federated learning portion of the experiment. The dataset was split into six parts, of which five represent the training set and are distributed to five different

linked nodes in such a manner that each node receives an adequate sample size from each of the categories, where test images are included in the sixth segment of the sequence.

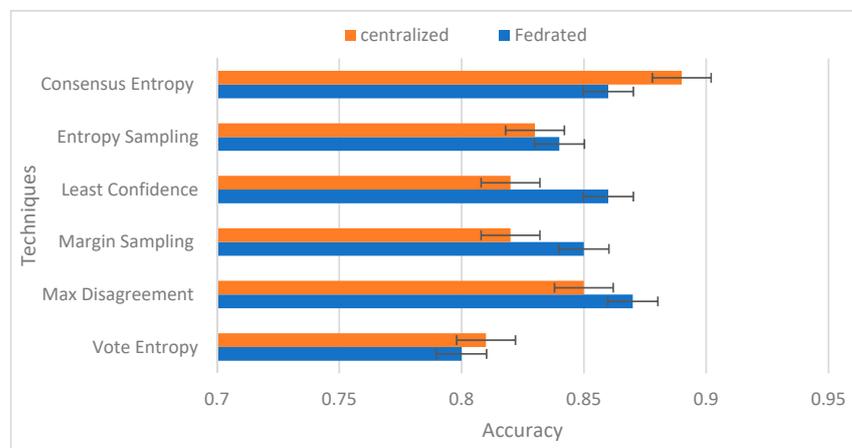
The LSTM learning model was used and is made up of four layers: a dropout layer; a classification layer; the first layer, which has 100 neurons; and the second layer, which contains 20 neurons. The data overfitting problem was resolved by introducing dropout and regularization techniques. This layer arbitrarily eliminated specific characteristics from the model by setting them to zero. The number of communication cycles and clients handled in each round within the FL framework was set to 50 and 5, respectively. This is another major characteristic of FL frameworks.

4.3. Experimental Results

The experimental results are shown in Figure 5, which depicts the results in terms of accuracy using the aforementioned AL techniques in an FL context. In the first 50 iterations, there is no significant difference in all six AL techniques, but afterwards, there is substantial variation in terms of accuracy utilizing these techniques. The performance of each approach stagnated after the 50th iteration. In particular, QBC approaches tend to acquire relatively fair higher accuracy rather than uncertainty approaches. The AL techniques are utilized to achieve better accuracy despite smaller sample sizes. Therefore, the QBC technique outperformed the uncertainty methods by achieving a relatively high accuracy despite the reduced training instances. In the majority of instances, with the exception of queries by the committee using consensus entropy and entropy-based disagreement strategies, the performance of FL is somewhat inferior to that of centralized learning.



(a)



(b)

Figure 5. Comparison of FL versus centralized learning contexts based on six participating nodes in various AL techniques. (a) Human sentiments; (b) Human physical activities.

The assessment of the above techniques is based on basic ML metrics, such as the F1-score, precision, recall, and test accuracy. These metrics are useful for assessing the above methods accurately, keeping in mind the imbalances in datasets in general and in implementing AL techniques. As shown in Tables 5 and 6, the values of the accuracy, recall, and F1-score follows a largely comparable pattern. Moreover, there is variation in these matrices, which motivated us to present the standard deviation among them to better understand the impact of AL in FL and centralized learning contexts.

Table 5. Comparative assessments of the AL approach in FL (client nodes = 6) and centralized learning contexts with respect to the weighted accuracy, recall, and F1-score under various sampling and disagreement techniques for human sentiments in the disaster situation dataset context.

Method	FL			Centralized Learning			Standard Deviation (FL–CL)		
	Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
Vote entropy	0.834	0.806	0.824	0.832	0.821	0.824	0.019116	0.023016	0.022417
Max disagreement	0.892	0.871	0.875	0.882	0.884	0.890	0.016384	0.018116	0.020189
Margin sampling	0.848	0.823	0.830	0.835	0.833	0.822	0.012209	0.014696	0.016822
Consensus entropy	0.874	0.865	0.870	0.856	0.861	0.852	0.009192	0.002828	0.006364
Entropy sampling	0.859	0.851	0.855	0.867	0.860	0.865	0.006752	0.005909	0.006344
Least confidence	0.866	0.842	0.850	0.856	0.852	0.856	0.006494	0.008424	0.007287

Table 6. Comparative assessments of the AL approach in FL (client nodes = 6) and centralized learning contexts with respect to the weighted accuracy, recall, and F1-score under various sampling and disagreement techniques for human physical activities in the disaster situation dataset context.

Method	FL			Centralized Learning			Standard Deviation (FL–CL)		
	Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
Vote entropy	0.853	0.846	0.831	0.821	0.825	0.823	0.020076	0.018964	0.019547
Max disagreement	0.882	0.881	0.879	0.876	0.881	0.860	0.013895	0.015545	0.017752
Margin sampling	0.864	0.851	0.840	0.845	0.851	0.818	0.015285	0.015212	0.017532
Consensus entropy	0.879	0.867	0.863	0.876	0.861	0.863	0.002121	0.018964	0.001414
Entropy sampling	0.881	0.875	0.846	0.897	0.890	0.875	0.009394	0.012553	0.011899
Least confidence	0.866	0.842	0.850	0.866	0.862	0.861	0.011502	0.015968	0.010289

4.4. Evaluation of the Impact on the Accuracy by the Number of Connected Nodes

Figure 6a,b illustrate the impact of the number of connected nodes on the accuracy, where there is a trade-off between the accuracy obtained from the aggregator model and the number of connected nodes in the FL context. The primary objective of this test is to determine whether extensions of connected nodes affect the accuracy of the federated models. It was evaluated with two nodes by distributing the dataset and then incrementally starting the connection of other nodes to participate. By connecting more nodes, the efficiency of both datasets experienced a considerable decrease, as shown in Figure 6. Initially, when there were fewer nodes, the accuracy in both datasets was consistent. However, as the number of nodes increased, the accuracy decreased more quickly, leading to a substantial decline in the number of training samples for each node. The drop in the accuracy of human sentiments in disaster data is somewhat greater than that of human physical activities in natural disaster scenarios. This indicates that there were fewer training instances in the dataset. The evaluation leads us to perform another experiment, in which we assess the effect of increasing the training set at each client through AL on the accuracy.

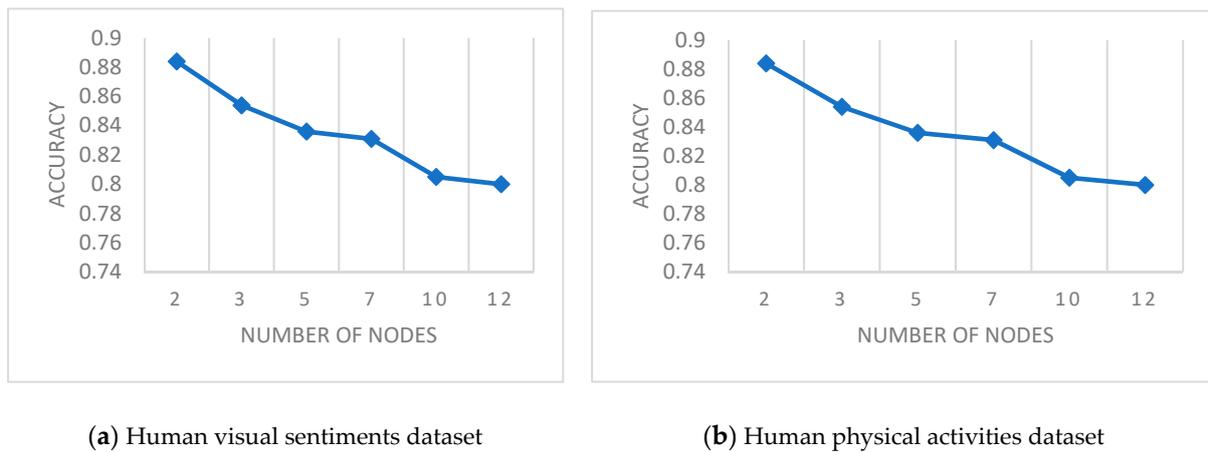


Figure 6. Impact on accuracy by expanding the number of participating client nodes in the FL environment: (a) human sentiment; (b) human physical activity.

4.5. Assessment of the Trade-off between the Training Data and Accuracy

Another experiment was performed, in which we distributed an increased number of training instances of both datasets with manually labeled instances to connected nodes to see how it impacts the accuracy in the FL environment by keeping the connected nodes constant. The training sets of both datasets contained a thousand images each, which were divided up and given to each of the five nodes to guarantee that each node would obtain an adequate quantity of training examples. We continued to expand the training set of each node by introducing an additional 50 images to each node’s dataset. This resulted in an increase of 230 training instances across all clients. Figure 7 demonstrates that the predictive accuracy improved when the number of training samples was increased. This was the case before and after the expansion. Early in the process, there was a great degree of diversity; however, as the number of training datasets obtained by each node grew, the pace of the improvement in their performance slowed down.

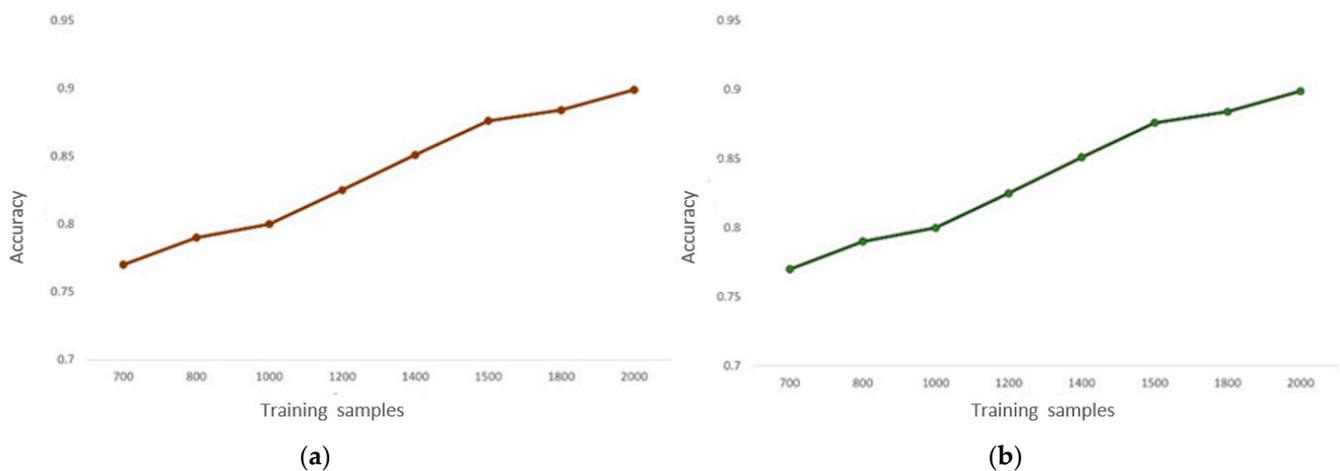


Figure 7. Accuracy and training sample size in FL is a trade-off during this experiment as there will only be a certain number of clients (i.e., six). As the number of training samples per client grows, so does the global model’s accuracy. (a) Human visual sentiments dataset; (b) Human physical activities dataset.

Tables 7 and 8 show the assessment/evaluation of the two baseline methods with regard to the level of accuracy and F1-score. (The Baseline 1 scenario is the ideal situation, in which each image is manually evaluated and labeled; the Baseline 2 scenario is a training set that contains unrelated data.) The suggested technique yields encouraging outcomes; on

both datasets, it performs better than Baseline 2, while maintaining a level of performance that is similar to Baseline 1.

Table 7. Accuracy and F1-score assessment versus the two starting points. Baseline 2 is training data containing unrelated images, Criteria 1 (manual label) is the best scenario, where each instance is carefully tagged and evaluated when compared to Baseline 2 and Baseline 1 across both datasets. The suggested technique achieves a significant improvement in human sentiments in the disaster situation dataset context.

Approaches	Centralized Learning		FL		Standard Deviation (FL–CL)	
	Accuracy	F1-Score	Accuracy	F1-Score	Accuracy	F1-Score
Criteria 1	0.731	0.742	0.710	0.724	0.096268	0.097621
Criteria 2	0.698	0.679	0.685	0.681	0.114799	0.120991
Proposed Approach	0.885	0.886	0.895	0.893	0.007071	0.00495

Table 8. Accuracy and F1-score assessment versus the two starting points. Baseline 2 is training data containing unrelated images, Criteria 1 (manual label) is the best scenario, where each instance is carefully tagged and evaluated when compared to Baseline 2 and Baseline 1 across both datasets. The suggested technique achieves a significant improvement in human physical activities in the disaster situation dataset context.

Approaches	Centralized Learning		FL		Standard Deviation (FL–CL)	
	Accuracy	F1-Score	Accuracy	F1-Score	Accuracy	F1-Score
Criteria 1	0.811	0.802	0.710	0.724	0.091884	0.08817
Criteria 2	0.856	0.847	0.685	0.681	0.100254	0.100323
Proposed Approach	0.891	0.888	0.899	0.896	0.005657	0.005657

4.6. Correlation against the Benchmarks

To demonstrate the efficacy of the AL techniques, Tables 6 and 7 show comparative results from using AL approaches to the two baselines in FL and a centralized learning context. For human sentiments in disaster situation datasets in federated and centralized learning, the model trained on the training samples labeled using the AL techniques and the manually labeled training sets yielded equivalent results. Nevertheless, the test accuracy remained higher than that indicated when trained models used training sets with AL techniques, which demonstrates the efficacy of AL techniques.

To understand the variations in the performance of the models in centralized and FL, we compare the standard deviation values. The lower scores in the standard deviation of the performance parameters using individual AL methods in centralized and FL environments indicate that FL has the potential to attain comparable accuracy with enhanced privacy.

There is a significantly greater difference between the performance of Criteria 1 and 2 and the suggested approach. Criteria 2 is the primary contributor to performance variance, because its results are lower in performance than those of Criteria 1 and the AL techniques.

Moreover, we compared our results with recent work that is conducted in AL-based FL environments, which are presented in Table 9. Aussel et al. [43] conducted their experiments on the MOA airlines dataset [44] and proposed a communication-efficient distributed learning approach based on active federated learning. They achieved a test accuracy of 61% with 10 participating clients and 30 communication rounds. Ahmed et al. [33] presented a similar approach to our approach where they used pre-trained ResNet [45] for feature extraction from two datasets. They analyzed five different active learning techniques in the federated learning environment. In their investigations, the performance of the techniques under different sampling and disagreement approaches varied significantly beyond the first

500 iterations. They attained an accuracy of 71% on a natural disaster image dataset using consensus entropy AL, and 82% on a waste classification dataset using consensus entropy AL in a federated learning environment. Ahn et al. [46] assessed annotation strategies when the AL algorithm is the maximum classifier discrepancy for AL (MCDAL) [47]. After learning auxiliary classifiers, it maximizes their prediction disparities. It substitutes conventional ambiguity using auxiliary classifier projections. This technique outperforms state-of-the-art AL algorithms on CIFAR-10 [48]. They create an FL framework in the annotation stage, where several active learning methodologies in the FL are contrasted, including traditional FL with client-level, separated active learning; active learning in federated active learning; and random sampling, where clients work together to perform the AL in order to choose the instances that are expected to be valuable to FL using a distributed optimization method. They achieved 51% accuracy with 49% recall and a 50% F1-score, respectively, using a CIFER-10 dataset. In our proposed approach where we analyzed six different AL techniques in the federated learning environment, we achieved comparable results to the centralized learning approach. These AL based methods include vote entropy, max. disagreement, margin sampling, consensus entropy, entropy and sampling, and least confidence. Tables 5 and 6 represent the acquired results for all six AL techniques which are achieved during experimentations in both federated and centralized learning contexts. We believe that comparable accuracy is achieved in the federated learning environment in the context of centralized learning.

Table 9. Comparison against benchmarks in AL-based FL methods in terms of accuracy, precision, recall and F1-score.

Author	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Aussel et al. [43]	0.61	0.60	0.58	0.51
Ahmed et al. [33]	0.71	0.70	0.70	0.73
Ahn et al. [46]	0.51	0.50	0.49	0.50
Proposed	0.89	0.88	0.87	0.84

5. Discussion

The experimental findings provide important takeaways:

- Compared with conventional passive learning, better results could be obtained with AL using fewer data instances.
- AL is equally favorable in centralized and FL contexts.
- Some approaches (or sampling/disagreement techniques) yield the maximum performance with fewer data than others. Therefore, when evaluating AL methods/query selection schemes, it is necessary to consider the amount of data needed to attain the best accuracy.
- It is essential to consider the stopping criteria for AL methods to maximize the effectiveness of utilizing these methods in various applications. This is because, after certain iterations, the algorithm is compelled to select instances that are non-pertinent, which can lead to a negative impact on the algorithm.
- The effectiveness of an algorithm is largely unaffected by FL. Nonetheless, a substantial number of training datasets are required at each client to fully train the local models. When there are insufficient training data available at each node to train the local model, the performance may be adversely impacted. In such instances, AL can be used to acquire relevant data images without human annotations.

6. Conclusions

In this article, we propose an AL-based FL framework that leverages unlabeled data at clients to train local models in two intriguing applications. A comprehensive analysis of two distinct pool-based AL techniques under various sampling and disagreement approaches

is presented. In addition, we demonstrate that AL may be appropriate in federated and centralized learning in general, and in applications that do not have access to extensively labeled datasets. Furthermore, we assessed the effect of a reduced node training dataset on the efficiency of the global model. In both datasets, we observed a considerable drop in performance. Initially, when there are fewer clients, the accuracy of both datasets is consistent; however, as the number of clients rises, the accuracy decreases more quickly, resulting in a considerable reduction in the number of training samples for each client. We handled AL as an offline process in the present implementation, such that it could automatically label training data at a client level before participating in FL.

In the future, we aim to expand the framework to active online learning. To achieve this, we will investigate the ways to integrate the unlabeled data in the training of local models throughout the many communication cycles of FL. This is a more difficult task for several reasons. In this scenario, the most significant difficulties would be twofold: if we retain the number of data instances to be selected at each iteration at a minimal value, then the number of communication cycles will increase; however, if we keep it at a large value, then there will be larger variances in the efficiency of the global model.

Author Contributions: Methodology, M.S.A.; Validation, W.-K.L.; Formal analysis, M.S.A.; Data curation, M.S.A.; Writing—original draft, M.S.A.; Writing—review & editing, W.-K.L.; Supervision, W.-K.L.; Project administration, W.-K.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Research Foundation of Korea (NRF) Grant funded by the Korean government (MSIT) (No. 2021R1A2C1014432).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The dataset is available on <https://tubo.tu.ac.kr/> (accessed on 2 January 2023).

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

List of Acronyms

Acronyms	Explanation
DNN	Deep neural networks
CNN	Convolutional neural network
FL	Federated learning
AL	Active learning
LSTM	Long short-term memory
IID	Independent identically distributed

References

- Li, T.; Sahu, A.K.; Talwalkar, A.; Smith, V. Federated Learning: Challenges, Methods, and Future Directions. *IEEE Signal Process. Mag.* **2020**, *37*, 50–60. [\[CrossRef\]](#)
- McMahan, H.B.; Moore, E.; Ramage, D.; Hampson, S.; y Arcas, B.A. Communication-Efficient Learning of Deep Networks from Decentralized Data. *Artif. Intell. Stat.* **2017**, *54*, 1273–1282.
- Li, T.; Sahu, A.K.; Zaheer, M.; Sanjabi, M.; Talwalkar, A.; Smith, V. Federated Optimization in Heterogeneous Networks. *arXiv* **2020**, arXiv:1812.06127.
- Smith, V.; Chiang, C.-K.; Sanjabi, M.; Talwalkar, A. Federated Multi-Task Learning. *arXiv* **2018**, arXiv:1705.10467.
- Bhowmick, A.; Duchi, J.; Freudiger, J.; Kapoor, G.; Rogers, R. Protection Against Reconstruction and Its Applications in Private Federated Learning. *arXiv* **2019**, arXiv:1812.00984.
- Bonawitz, K.; Ivanov, V.; Kreuter, B.; Marcedone, A.; McMahan, H.B.; Patel, S.; Ramage, D.; Segal, A.; Seth, K. Practical Secure Aggregation for Federated Learning on User-Held Data. *arXiv* **2016**, arXiv:1611.04482.
- Bonawitz, K.; Eichner, H.; Grieskamp, W.; Huba, D.; Ingerman, A.; Ivanov, V.; Kiddon, C.; Konečný, J.; Mazzocchi, S.; McMahan, H.B.; et al. Towards Federated Learning at Scale: System Design. *Proc. Mach. Learn. Syst.* **2019**, *1*, 374–388.
- Kairouz, P.; McMahan, H.B.; Avent, B.; Bellet, A.; Bennis, M.; Bhagoji, A.N.; Bonawitz, K.; Charles, Z.; Cormode, G.; Cummings, R.; et al. Advances and Open Problems in Federated Learning. *Found. Trends® Mach. Learn.* **2021**, *14*, 1–210. [\[CrossRef\]](#)

9. Lim, W.Y.B.; Luong, N.C.; Hoang, D.T.; Jiao, Y.; Liang, Y.-C.; Yang, Q.; Niyato, D.; Miao, C. Federated Learning in Mobile Edge Networks: A Comprehensive Survey. *IEEE Commun. Surv. Tutor.* **2020**, *22*, 2031–2063. [[CrossRef](#)]
10. Tashu, T.M.; Hajiyeva, S.; Horvath, T. Multimodal Emotion Recognition from Art Using Sequential Co-Attention. *J. Imaging* **2021**, *7*, 157. [[CrossRef](#)]
11. Sun, Y.; Loparo, K. Context Aware Image Annotation in Active Learning. *arXiv* **2020**, arXiv:2002.02775.
12. Liu, W.; Chang, X.; Chen, L.; Phung, D.; Zhang, X.; Yang, Y.; Hauptmann, A.G. Pair-Based Uncertainty and Diversity Promoting Early Active Learning for Person Re-Identification. *ACM Trans. Intell. Syst. Technol.* **2020**, *21*, 21. [[CrossRef](#)]
13. Ngo, G.T.; Ngo, T.Q.; Nguyen, D.D. Image Retrieval with Relevance Feedback Using SVM Active Learning. *Int. J. Electr. Comput. Eng.* **2016**, *6*, 3238–3246. [[CrossRef](#)]
14. Yuan, J.; Hou, X.; Xiao, Y.; Cao, D.; Guan, W.; Nie, L. Multi-Criteria Active Deep Learning for Image Classification. *Knowl. Based Syst.* **2019**, *172*, 86–94. [[CrossRef](#)]
15. Cao, X.; Yao, J.; Xu, Z.; Meng, D. Hyperspectral Image Classification With Convolutional Neural Network and Active Learning. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 4604–4616. [[CrossRef](#)]
16. Mu, C.; Liu, J.; Liu, Y.; Liu, Y. Hyperspectral Image Classification Based on Active Learning and Spectral-Spatial Feature Fusion Using Spatial Coordinates. *IEEE Access* **2020**, *8*, 6768–6781. [[CrossRef](#)]
17. Li, X.; Liu, B.; Zhang, K.; Chen, H.; Cao, W.; Liu, W.; Tao, D. Multi-View Learning for Hyperspectral Image Classification: An Overview. *Neurocomputing* **2022**, *500*, 499–517. [[CrossRef](#)]
18. Ortis, A.; Farinella, G.M.; Battiato, S. Survey on Visual Sentiment Analysis. *IET Image Process.* **2020**, *14*, 1440–1456. [[CrossRef](#)]
19. Zhu, Y.; Zhu, Y.; Ge, N.; Gao, W.; Zhang, W. Visual Emotion Analysis via Affective Semantic Concept Discovery. *Sci. Program.* **2022**, *2022*, e6975490. [[CrossRef](#)]
20. Priya, D.T.; Udayan, J.D. Affective Emotion Classification Using Feature Vector of Image Based on Visual Concepts. *Int. J. Electr. Eng. Educ.* **2020**. [[CrossRef](#)]
21. Jou, B.; Chen, T.; Pappas, N.; Redi, M.; Topkara, M.; Chang, S.-F. Visual Affect Around the World: A Large-Scale Multilingual Visual Sentiment Ontology. In Proceedings of the 23rd ACM International Conference on Multimedia, Brisbane, Australia, 26–30 October 2015; pp. 159–168.
22. Sadiq, A.M.; Ahn, H.; Choi, Y.B. Human Sentiment and Activity Recognition in Disaster Situations Using Social Media Images Based on Deep Learning. *Sensors* **2020**, *20*, 7115. [[CrossRef](#)] [[PubMed](#)]
23. Borth, D.; Ji, R.; Chen, T.; Breuel, T.; Chang, S.-F. Large-Scale Visual Sentiment Ontology and Detectors Using Adjective Noun Pairs. In Proceedings of the 21st ACM International Conference on Multimedia, Barcelona, Spain, 21–25 October 2013; Association for Computing Machinery: New York, NY, USA, 2013; pp. 223–232.
24. Chen, T.; Borth, D.; Darrell, T.; Chang, S.-F. DeepSentiBank: Visual Sentiment Concept Classification with Deep Convolutional Neural Networks. *arXiv* **2014**, arXiv:1410.8586.
25. Al-Halah, Z.; Aitken, A.; Shi, W.; Caballero, J. Smile, Be Happy :) Emoji Embedding for Visual Sentiment Analysis. *arXiv* **2020**, arXiv:1907.06160.
26. Liu, C.; Fang, F.; Lin, X.; Cai, T.; Tan, X.; Liu, J.; Lu, X. Improving Sentiment Analysis Accuracy with Emoji Embedding. *J. Saf. Sci. Resil.* **2021**, *2*, 246–252. [[CrossRef](#)]
27. Huang, F.; Wei, K.; Weng, J.; Li, Z. Attention-Based Modality-Gated Networks for Image-Text Sentiment Analysis. *ACM Trans. Multimed. Comput. Commun. Appl.* **2020**, *16*, 1–19. [[CrossRef](#)]
28. Memiş, G.; Sert, M. Detection of Basic Human Physical Activities With Indoor–Outdoor Information Using Sigma-Based Features and Deep Learning. *IEEE Sens. J.* **2019**, *19*, 7565–7574. [[CrossRef](#)]
29. An, S.; Bhat, G.; Gumussoy, S.; Ogras, U. Transfer Learning for Human Activity Recognition Using Representational Analysis of Neural Networks. *arXiv* **2021**, arXiv:2012.04479. [[CrossRef](#)]
30. Amin, M.S.; Yasir, S.M.; Ahn, H. Recognition of Pashto Handwritten Characters Based on Deep Learning. *Sensors* **2020**, *20*, 5884. [[CrossRef](#)]
31. Amin, M.S.; Ahn, H. Earthquake Disaster Avoidance Learning System Using Deep Learning. *Cogn. Syst. Res.* **2021**, *66*, 221–235. [[CrossRef](#)]
32. Braun, M.; Krebs, S.; Flohr, F.; Gavrilu, D.M. EuroCity Persons: A Novel Benchmark for Person Detection in Traffic Scenes. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *41*, 1844–1861. [[CrossRef](#)]
33. Ahmed, L.; Ahmad, K.; Said, N.; Qolomany, B.; Qadir, J.; Al-Fuqaha, A. Active Learning Based Federated Learning for Waste and Natural Disaster Image Classification. *IEEE Access* **2020**, *8*, 208518–208531. [[CrossRef](#)]
34. Sun, T.; Li, D.; Wang, B. Decentralized Federated Averaging. *arXiv* **2021**, arXiv:2104.11375. [[CrossRef](#)] [[PubMed](#)]
35. Zhu, H.; Jin, Y. Multi-Objective Evolutionary Federated Learning. *arXiv* **2019**, arXiv:1812.07478. [[CrossRef](#)] [[PubMed](#)]
36. Cho, K.; van Merriënboer, B.; Gulcehre, C.; Bahdanau, D.; Bougares, F.; Schwenk, H.; Bengio, Y. Learning Phrase Representations Using RNN Encoder-Decoder for Statistical Machine Translation. *arXiv* **2014**, arXiv:1406.1078.
37. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [[CrossRef](#)]
38. Guo, Y.; Liu, Y.; Bakker, E.M.; Guo, Y.; Lew, M.S. CNN-RNN: A Large-Scale Hierarchical Image Classification Framework. *Multimed. Tools Appl.* **2018**, *77*, 10251–10271. [[CrossRef](#)]
39. Soleymani, M.; Garcia, D.; Jou, B.; Schuller, B.; Chang, S.-F.; Pantic, M. A Survey of Multimodal Sentiment Analysis. *Image Vis. Comput.* **2017**, *65*, 3–14. [[CrossRef](#)]

40. Cowen, A.S.; Keltner, D. Self-Report Captures 27 Distinct Categories of Emotion Bridged by Continuous Gradients. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, E7900–E7909. [CrossRef] [PubMed]
41. Cognitive Robotics Lab. Industry University Cooperation Education Hub. Tongmyong University. Available online: <https://tubo.tu.ac.kr/> (accessed on 11 February 2023).
42. Find the Best Global Talent. Available online: <https://www.fiverr.com/> (accessed on 11 February 2023).
43. Aussel, N.; Chabridon, S.; Petetin, Y. Combining Federated and Active Learning for Communication-Efficient Distributed Failure Prediction in Aeronautics. *arXiv* **2020**, arXiv:2001.07504.
44. Download MOA—Massive Online Analysis from SourceForge. Available online: <https://sourceforge.net/projects/moa-datastream/files/Datasets/Classification/airlines.arff.zip/download> (accessed on 22 December 2022).
45. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. *arXiv* **2015**, arXiv:1512.03385.
46. Ahn, J.-H.; Kim, K.; Koh, J.; Li, Q. Federated Active Learning (F-AL): An Efficient Annotation Strategy for Federated Learning. *arXiv* **2022**, arXiv:2202.00195.
47. Cho, J.W.; Kim, D.-J.; Jung, Y.; Kweon, I.S. MCDAL: Maximum Classifier Discrepancy for Active Learning. *IEEE Trans. Neural Netw. Learn. Syst.* **2022**, 1–11. [CrossRef] [PubMed]
48. Krizhevsky, A. *Convolutional Deep Belief Networks on CIFAR-10*; University of Toronto: Toronto, ON, Canada, 2009.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.