Int. j. inf. tecnol.
https://doi.org/10.1007/s41870-023-01307-6

ORIGINAL RESEARCH

Emergency events detection based on integration of federated learning and active learning

Khalid Alfalqi[1] · **Martine Bellaiche**[1]

Received: 12 November 2022 / Accepted: 15 May 2023

Abstract Social media networks now make it easy to access, in real-time, massive amounts of information from all over the world. They are often the primary source of information for billions of people. In addition to other uses, social media platforms can be valuable to provide real-time information about emergency events and can help detect emergency situations. This can provide affected individuals with crucial information about imminent dangers, prompting them to take appropriate action and significantly influencing their decisions. The continuous flow of data generated during these emergencies requires efficient training and modeling. Deep learning can simplify this process and make it less reliant on feature extraction strategies. However, implementing deep learning-based solutions can be expensive and time-consuming due to the vast amounts of data involved. To address these limitations, in this paper, we employed and analyzed the effectiveness of utilizing Active Learning techniques along with Federated Learning for emergency events using a dataset garnered from the social media. After collecting images, we used a Federated Learning paradigm to split the data amongst different clients and used vision transformers as local models for each client. Furthermore, we used an ensemble approach to integrate the results from these strengthened local and global models in a novel way, demonstrating that the proposed setup yields an accuracy of 99%. The Logloss score reached 0.043, indicating the outstanding performance of our approach.

✉ Martine Bellaiche
martine.bellaiche@polymtl.ca

Khalid Alfalqi
khalid.alfalqi@polymtl.ca

[1] Computer and Software Engineering Department, Ecole Polytechnique, Montreal, Canada

Keywords Federated learning · Active learning · Social networks · Object detection · Emergency event detection · Twitter · Snapchat · Vision transformers ViT

1 Introduction

The abundant growth and widespread use of social media across the globe provide faster and more efficient ways to communicate, exchange information and share data. This phenomenon totally transformed the world in which we live in today. These networks provide billions of people the opportunity to join a single platform to exchange interests, share information, multi-media contact, news and other items of interest with their friends, contacts as well as the general public. Aside from these common scenarios, social media networks can also be used for timely and prompt detection of emergency events. This would be beneficial for everyone in the proximity of the event as they could be provided with useful information indicating imminent dangers thereby influencing their real-time decision-making to take appropriate action.

An emergency is any untoward situation that poses a direct potential risk to human lives, property and infrastructure although the associated dangers are still within control [7]. Fires, floods, earthquakes, train and aircraft accidents, chemical spills, radiation exposure, threats of personal violence, acts of terrorism, and public health crises such as epidemics and pandemics are some examples of these emergencies [12]. However, when the impact of the emergency on humans, materials, economics or the environment is beyond the control of the affected community, it becomes a disaster [30].

The detection of these emergency events has significantly been aided by analytical information obtained from

Published online: 27 June 2023

social media platforms [21]. Therefore, it is very important to communicate essential information about the emergency situation to the response teams as soon as a crisis emerges. In this regard, deep learning-based models were recently created to classify these emergency events based on the data provided by social media [36]. The fundamental benefit of using deep learning to classify and detect problems is that the entire network is trained—from raw data to final classification results, which decreases the requirements to build an appropriate feature extraction strategy [29]. Moreover, it requires a large number of high-quality samples for model training. Since the data labelling procedure is particularly expensive and time-consuming in the social networks dataset due to the large amount of data available on such platforms, which can limit the implementation of deep learning technologies. In [31], Simon et al have described in detail how significant is the rule of social media in engaging the citizens for both, the dissemination as well as access to information in the events of disasters such as earthquake, a hurricane or even a terrorist attack. Similarly, in another review paper [22], Martinez et al performed a systematic review of literature, which aims to present an outline of the present research status on the utilization of Twitter in emergency management. Additionally, the study highlights the obstacles and potential research avenues in this area.

Lately, Federated Learning (FL) techniques have also been employed to facilitate the development of a shared prediction model while retaining individual training data on the device [23]. Here, each user develops a local training model by maximizing its optimal solution, which is exchanged with the aggregator that combines them to generate a global model that is established and distributed to all network users. Individual users then adjust their own model according to what they have learned from the global model. Finally, the FL client communication procedure is repeated until the global model obtains sufficient precision [39]. In this way, privacy and bandwidth-related limitations are aptly handled by the use of FL as the training data is shared by the trained models rather than each individual node [40].

FL is a decentralized machine learning approach that enables network users to train using their localized data rather than exchanging relevant data with the global server.

In addition to these techniques, Active Learning (AL) is a way to decrease the amount of manual labelling that must be done for learning during the disaster, thereby eliminating the need for human interference or examination of training samples. This technique also helps to resolve ambiguities in the learning process by interacting with a human operator during the learning process. In situations where few data points are available or in contexts where collecting or labeling new information is expensive, AL is a useful approach. As a result of AL, samples are selected in ways that minimize sample numbers while maximizing model efficiency [38].

To further optimize the problem of emergency detection using social media data, ensemble-based approaches have also been applied. An ensemble approach combines predictions from several models that fit the same data and is intended to achieve better results. In this process, models are created for use in the ensemble and assumptions are made to predict how all members of the ensemble will be combined [6]. Moreover, transfer learning techniques have also been used where a model is initially trained on one task and, later, some or all of it is used to begin learning about an additional task. A related task with a large amount of data can be used in conjunction with the main task of interest. As models are incrementally trained, transfer learning can be useful to continue to train them [37].

Vision Transformers (ViT) was proposed in 2017 by Vaswani et al. as a sequence-to-sequence (Seq2Seq) model. Its main innovation is that it can learn from the most informative elements by entirely relying on the attention mechanism. By applying an attention mechanism to connect foreign words, the Transformer proposes a new approach that encodes each position, enabling parallelization, thereby accelerating learning. The mechanism is based on self-attention [11]. The rationale of the proposed research work is further highlighted below:

- In emergency situations, support and rescue services need to react quickly and precisely. A continuous flow of data is provided by the social networks, but most of it is heterogeneous and requires further labelling before being used in near-real-time applications. However, the manual labelling process is time-consuming and may not be rapid enough to allow efficient information utilization. Therefore, using the vision transformer model will speed up the model.
- The redundant training of models is a significant concern for emergency event detection. Traditional deep learning models need to train the whole dataset to be able to produce sufficient predictions. To be able to provide near real-time predictions to first responders, the model must query the latest data associated with the incidents as time progresses to keep up with the latest developments. The active learning approach may therefore be able to adapt to such rapidly changing tasks.
- Instead of using a central model that collects data from different locations—which causes latency and increases the response time to detect emergency events—applying the FL approach will help distribute the models among clients that can train the model locally and then aggregate it in a central server, which is a more viable solution.

The proposed work presents the following research contributions:

1. A decentralized and collaborative framework is proposed: it incorporates FL and AL as well as the ViT model.
2. The training at client machines is boosted using an AL-based framework, which selects a subset of the whole dataset to accelerate the training process.
3. It proposes a combination of global and local models, weights average probabilities to construct a weighted average ensemble model and evaluates the performance of the approach.

The remainder of the paper is organized as follows: Sect. 2 describes the existing research work carried out on emergency events using federated and AL techniques, followed by the problem definition. Section 3 depicts the research methodology phases, as well as the approach and models employed in the experiments. Section 4 addresses the experimental environment construction, the evaluation measures, the performance analysis for our suggested approach, and a discussion regarding the outcomes. Finally, Sect. 5 highlights the findings and suggests future study avenues.

2 Related work

Many existing research works have used Federated Learning, Active Learning and Vision Transformer approaches to detect emergency events but limited work has been conducted on the subject with regards to social media networks. In this section, we present the existing literature that has used FL, AL and ViT techniques to detect emergency events.

2.1 Federated learning to detect emergency events

Zhang et al. [40] proposed an FL approach to sort disasters in common computation systems. To maintain the confidentiality of the data stored in the social computing nodes in the federated environment, Paillier homomorphic encryption technique is used. Specifically, the FL system uses transfer learning technology to reduce computation and communication costs. The FL-TL approach is validated using an actual catastrophe picture collection compiled from several social networks. The findings from both theoretical examinations and experimental trials demonstrate that the FL-TL is a practical, secure, flexible and efficient technology and that can be integrated easily into different transferable learning methods. The proposed work yields a 73% accuracy compared with other research works.

Mittal et al. [25] presented a framework that maps Social Media Postings (SMPs) like tweets and posts to appropriate tags using Social Media Engines. This technology is unique as it combines namespace integration with federated active learning and interpretation techniques to quickly identify and send critical SMPs to the appropriate first responders in a dispersed multi-organizational context. The proposed technique outperforms a straightforward keyword-based classification and various other NLP-based categorization models on real-world data (including tweets sent by people during the California wildfires in 2018).

Li et al. [20] proposed a unique FL method, known as Hybrid Federated Learning (HFL), intended to attain a learning equilibrium in terms of efficiency and effectiveness. This was done to reduce the impact that stragglers impose on the overall process. An adaptive approximation approach is known as Adaptive Delayed-Stochastic Gradient Descent (AD-SGD) that combines the delayed local updates with the joint model was suggested. By using this technique, the achieved accuracy resulted were 12% higher.

Priyanshu et al. [26] proposed a federated averaging method. The model should be consolidated for continuous learning using tensor flow transformation by employing distributed learning. Overall, a 76.4% accuracy result was achieved through the federated averaging method.

2.2 Active learning usage of emergency events applications

Active learning is a form of machine learning technique which can be used to categorize data with intended results. Said et al. [28] examined the effectiveness of using active learning algorithms to analyze disaster-related photos from the social media and assessed their performance in terms of categorization accuracy. They have used pool-based sampling and Support Vector Machine (SVM) classifier transformation. The pool-based sampling method involves image samples from a pool of unlabeled pictures and incorporates them into an early small labeled training dataset. Promising results were obtained with F-scores of 0.727 and 0.670.

Ahmed et al. [2] proposed a federated Active Learning-based framework by implementing and analyzing a variety of AL approaches in two distinct application areas. It was shown that the proposed technique was independent of the dataset or application that was being used by assessing the proposed method in two intriguing applications—Natural Catastrophe Analysis and Waste Classification—which have distinct characteristics and obstacles. Both applications demonstrated promising results, with 3.1% and 4% improvements in accuracy over the training sets.

In another similar work, Bugel et al. [5] analyzed multilingual tweets, particularly focusing on Mediterranean languages, such as Greek, Turkish and Romanian, to detect early warnings for earthquakes and tsunamis. They emphasized that most of such events are described in the poster's native languages on Twitter, hence the importance of developing frameworks that can support multiple languages.

Wang and Brenning [34] proposed an Active Learning-based technique to map landslides using SVM. It was observed that the performance of models trained with randomly collected data was worse than the ones trained using active-learning methodologies to detect landslide emergencies. With only 285 examples, the mean AUROCs of the SVM using margin sampling as an active-learning approach was 0.80, but it increased to 0.83 with a set of 410 examples. On the other hand, SVMs that used query-by-committee and random sampling could attain AUROCs in the vicinity of 0.79, although this was only the case for very high sample sizes. Meanwhile, the SVM that included margin sampling proved to be the most reliable of the techniques. YOLO object detection method was used by Bommel [4] by employing deep CNN transformation and advanced machine learning algorithms. A high precision level on homogeneous data was achieved using the YOLO object detection method.

In [27], Ragini et al. proposed a model that collects disaster data from the social networks. It categorizes them based on the needs of those affected and uses Active Learning to analyze the data sentiment. The study concludes that a lexicon-based approach is the most effective method to analyze people's needs during disasters. The practical implications of this methodology include real-time categorization and classification of social media big data for disaster response and recovery, aiding emergency responders and rescue staff in developing better strategies to manage rapidly changing disaster environments. ANN was used to perform classification.

2.3 Vision transformers application for emergency events

Horváth et al. [13] proposed an unsupervised method that uses (ViT) to detect manipulations in satellite images. The proposed research work used splicing detection along with unsupervised learning to accomplish the desired results. Morphological filters were the primary tools to apply vision transformations. With an F-score of 0.364, the system's efficiency increased significantly.

Liu et al. [33] performed a study on ViT for the purpose of extracting data pertaining to buildings. Both the global context route and the detailed context path were used during this study. A design with a global context route and a spatial-detailed context path was used to encode rich geographical features and capture global relationships. The most cutting-edge outcomes, in terms of performance, were attained at a rate of 75.74 %.

Kaur et al. [16] proposed a network that uses hierarchical spatial features and captures temporal differences to achieve state-of-the-art performance for building localization, damage classification, and change detection tasks on large-scale disaster damage data sets. The work also introduces a new high-resolution satellite imagery data set and demonstrates an approach that uses it for domain adaptation with limited fine-tuning.

Another study, using the asymmetric twin network approach by Xia et al. [35] employed self-supervised comparison learning to obtain remarkable results with substantial potential and generated an F-score of 0.678. Guo et al. [10] modelled the effects of natural hazards and provided decision-making assistance. The research work presented an automated approach to extract building information from satellites and street view photos and it applied a unique transformer-based deep neural network technique. A multi-domain learning technique was used to construct a single compact model for several image-based supervised neural knowledge abstraction tasks using numerous data sources. The F-score obtained after employing this technique was 0.89% with 93.54% accuracy.

Da et al. [8] used a two-staged damage assessment method by employing symmetrical hierarchical transformation using deep learning which outperformed CNN-based frameworks in extracting long-range semantic data to estimate the damage. This method yielded an 80% accuracy level.

Table 1 shows the significant accuracy of the results obtained through Federated and Active Learning. That depicts the significance of both approaches, and ViT has been widely implemented in applications for disasters in social networks.

2.4 Analysis of existing techniques

Many research works have employed Artificial Intelligence (AI) based techniques which use social media data/information for emergency detection.

Research works mentioned in Sect. 2.1 propose ambitious perspectives to detect emergency events by applying FL to the collected dataset from social networks and other sources. However, none of the aforementioned papers have employed the AL and ViT.

Moreover, research works in Sect. 2.2 have employed AL approaches to solve the limitation of training large datasets but none of the research works consider using the ViT model as a reliable image processing method that can replace both traditional CNN image processing and object detection methods.

Moreover, ViT has been utilized in the research works discussed in Sect. 2.3 but they are based on traditional deep learning approaches. Moreover, their dataset was collected in predefined locations. In contrast, our approach relies on the datasets collected from Snapchat map API and Twitter API to find the precise location of emergency events.

Table 1 A comparison and summary of the existing literature

References	Technique used	Results	AL component	Transformer used
[40]	Paillier homomorphic encryption method	State-of-the-art performance 75.74% accuracy	No	Yes
[33]	Global context path spatial-detailed context path	Efficiency increased up to 73%	No	No
[35]	Asymmetric twin network method	F-score: 0.678	No	Yes
[2]	Active learning-based federated method	Results, improvement of 3.1% and 4%	Yes	No
[13]	Splicing detection method	–	No	Yes
[8]	Two-stage damage assessment (SDA former) method	Significant improvement in the large-scale building damage	No	Yes
[28]	Pool-based sampling	F-score: 0.727	Yes	No
[10]	Multi domain learning	Accuracy: 93.54%	No	Yes
[20]	Adaptive delayed-SGD	–	No	No
[4]	YOLO-object detection method	High precision level on homogeneous data	Yes	No
[26]	Federated averaging method	Accuracy: 80.52%	No	No
[34]	Uncertainty sampling and query method	Accuracy: 80.52%	Yes	No
[25]	FLARE, a framework using SMEs and SMPs	Promising results as compared to other strategies	Yes	No
[16]	Damage classification—temporal change detection	F-score: 0.796 (damage classification) F-score: 0.908 (change detection)	Yes	Yes
[27]	Disaster text classification using BOW and POS features	F-score approaching 97% in various categories	Yes	No

3 Proposed framework for emergency events detection

This section presents the complete proposed framework, comprising all modules that are integrated together to achieve the goal of disaster event detection. The process starts with the collection of a dataset from the Snapchat map API and Twitter API in the exact locations of the emergency events that were identified earlier. In the second step, the dataset is split and distributed across all the clients in order to demonstrate the working of federated learning. In the third step, data pre-processing and data augmentations methods are employed to enhance the model training. It is followed by the creation of ViT and executed on all clients, as the fourth step. In the fifth step, a global model is configured by aggregating the local models, trained on each model. This aggregated model is then shared with the all the clients. Finally, an ensemble learning approach is deployed to further enhance the performance at the client machines. The overall framework is presented in Fig. 1.

The steps of the proposed model are shown in Fig. 2.

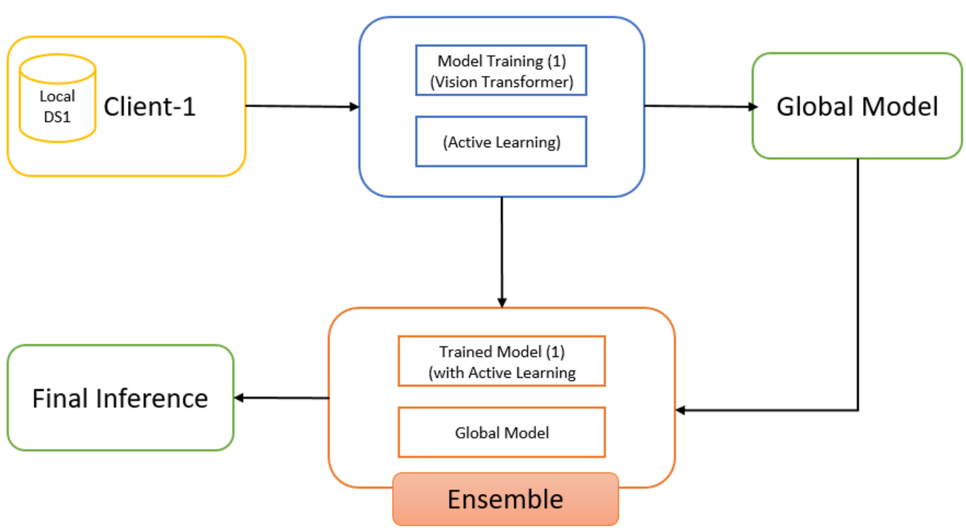

Fig. 1 The framework (with perspective of a single client)

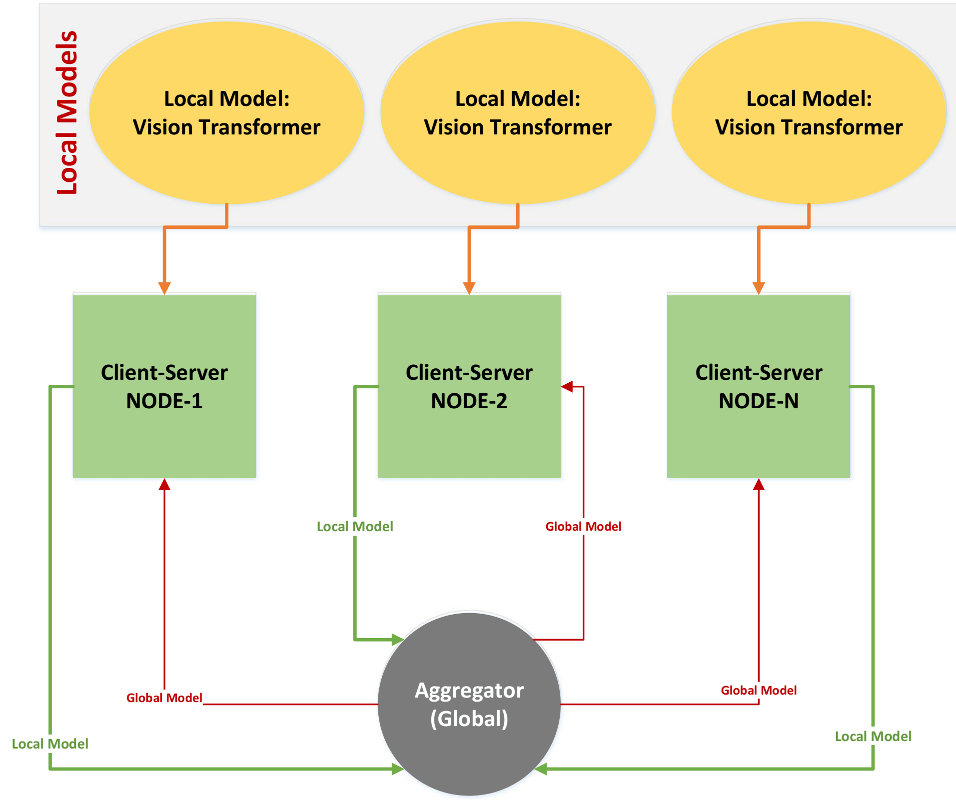

Fig. 2 The federated learning setup

3.1 Data collection and preparation

A dataset comprises images collected from the locations where the disasters occurred, using Snapchat. For this purpose, three hot-spot zones in the Snapchat map are identified from a six-month period, from March to August 2022. The dataset is collected using Node.js wrapper, that we created in our previous work [3]. After identifying the exact location of these emergency events, the dataset is collected using the Snapchat API, from the location where the emergency events occurred. More particularly, we collected data from three following events:

- The fire outbreak that occurred in Delhi, India, on May 13, 2022.
- The flood snaps collected from the flood in Eastern Cape in South Africa on April 11, 2022.
- The earthquake snaps taken from Fiji in Japan, in August 2022.

The complete detail of the collection of datasets is presented in [3]. In addition to the existing dataset, collected from the Snapchat API, we enriched our dataset by leveraging the Twitter API to collect more images related to these events and their locations.

The dataset used in this research work is composed of 4052 images with three categories of emergency events: earthquake, flood and fire. The dataset is split into training and testing with the 80:20 split and stratified based on the target distribution in order to keep the same distribution over both sets. We wanted this split to be random to avoid any bias to the testing set and generate fair results. The dataset is further split into multiple folds, depending on the number of clients configured for Federated Learning. This was done using the Sklearn method called *StratifiedKFold* which guarantees equal splits for the dataset over clients with the same target distribution over all of them. The statistics of the dataset are presented in Table 2

3.2 Processing performed on each client

The following steps, comprising the data pre-processing, feature extraction, model configuration and model training, are performed on each client separately.

Table 2 Statistics of the data collected from Snapchat and Twitter

	Snapchat	Twitter
Total images	3012	1040
Earthquake	1050	350
Fire	1032	240
Flood	930	450

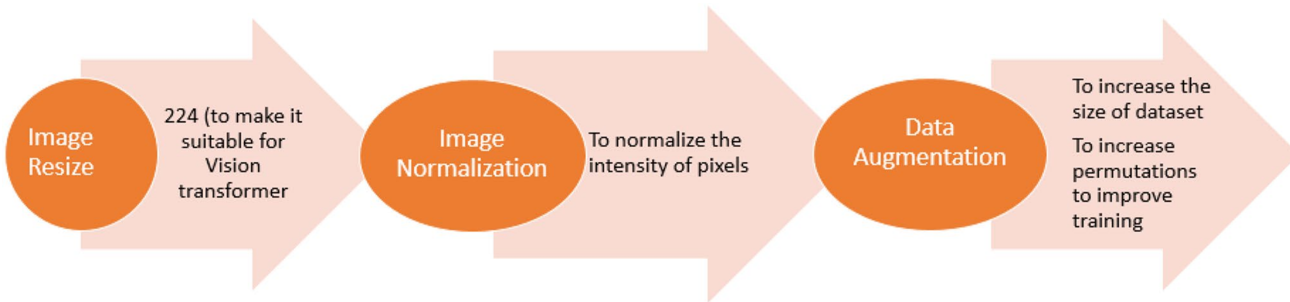

Fig. 3 The data pre-processing steps

3.2.1 Data pre-processing

During the first phase of pre-processing, images are resized as ViTs require the dataset to be scaled to a consistent size of the same dimensions. We resized the collected images to a length of *224*. Then, we applied image normalization, which is a method to alter the spread of pixel intensity. Its main function is to normalize an incoming picture towards a set of adjacent pixels that are increasingly recognizable or typically perceived. Subsequently, the dataset is augmented by creating different replicas of the original images. It is performed to expand the dataset and introduce the neural networks to a broad range of image permutations [32]. In addition, it will increase the likelihood that the model will detect objects in any configuration. Data augmentation may successfully prevent fitting problems throughout the advanced training phase and it can tremendously enhance data clarity [17]. Several data augmentation methods were employed, including vertical and horizontal flipping, rotating to a specific angle (less than 20°) and raising or reducing brightness. Figure 3 shows the data pre-processing steps.

Table 3 Hyper-parameters of the vision transformer model

Models	Hyper-parameters
ViT	TRAIN_SIZE = 0.8
	IMG_SIZE = 224
	N_NODES = 2
	N_LABELED_POINTS = 500
	AL_POINTS = 200
	BS = 32
	GLOBAL_N_EPOCHS = 10
	LOCAL_N_EPOCHS = 3
	MODEL = "beit_base_patch16_224_in22k"
	PRETRAINED = True
	CRITERION = nn.CrossEntropyLoss()
	SKF = StratifiedKFold(n_splits=5)
	ADAM OPTIMIZER LR = 1E-4
Federated learning	N_NODES: 2 GLOBAL_N_EPOCHS: 10
Active learning	N_LABELED_POINTS: 500 (initial points to train on the active learning)
	AL_POINTS: 200 (maximum Points queried using the Least confidence sampling strategy)
	LOCAL_N_EPOCHS: 3 (number of epochs during active learning)
Pre-trained modelL	“beit_base_patch_16_224_in22k”
	BeiT base patch 16 image_size224
	pretrained on ImageNet-22k
	Batch_size: 32
	Learning_rate: 1e-4
	Img_size: 224x224
	Scheduler: ReduceLROnPlateau with params
	factor=0.5,patience=1,min_lr=1e6,mode="min"
The proposed ensemble	Estimators = Global model and local model, weighted average ensembling
	Voting = soft

3.2.2 Feature extraction

After completing the data pre-processing, the next step consists of performing feature extraction, which is conducted by following these steps:

- Create patches from an image;
- The patches need to be flattened;
- Using the flattened patches, create linear embeddings with a lower dimension;
- Positional embeddings are added;
- Using a transformer encoder, feed the sequence as an input.

3.2.3 Creation of vision transformer model

We then created and trained a ViT model on our dataset using the transfer learning approach. We created our ViT by utilizing the pre-trained model beit_base_transformer with patch_size = 16 and img_size = 224. We loaded the transformer model's pre-trained weights on Imagenet22k. Table 3 shows the hyper-parameter of the model. It should be noted that, each client trains the ViT Model on its own local dataset.

3.3 Updating the global model using the federated learning approach

Algorithm 1 Active Learning in Federated Learning for the Transformer Model

Input:
E: Global Number of Epochs,
E': Local Number of Epochs,
GM: GlobalModel,
N: Number of Clients C_k where $k = 1..N$
US_k: Unlabelled set of images
LS_k: Labeled images,
NAL: Maximum queried active learning samples

Output:
W_{GM}: Global model weights with average weights from all clients

Load GM weights using transfer learning and save as W_{GM}
Initialize k Clients C_k
    $loss_k = infinity$
    $bestw_k = W_{GM}$

for each global epoch from 1 to **E do**
    for each client k from 1 to **N do**
        $LS_k^{train}, LS_k^{valid} \leftarrow Split(LS_k)$
        Client C_k load W_{GM}
        Local_Training(LS_k^{train})
        for each local epoch from 1 to **E' do**
            USLC $\leftarrow ClientC_k$ does inference on US_k
            Get NAL samples according to least confidence sampling strategy.
            $US_k \leftarrow US \setminus USLC$
            $LS_k^{train} \leftarrow LS_k^{train} \cup USLC$ labeled.
            Client C_k does local training on LS_k^{train}
            $eval\_loss \leftarrow Calculate\_Validation\_Loss()$
            if$(eval\_loss < loss_k)$
            $loss_k \leftarrow eval\_loss$
            $bestw_k \leftarrow W_k^{e'}$
            endif
        end for
    end for
    $W_{GM} = (1/\mathrm{N}) \times \sum_{k=1}^{N} bestw_k$
end for
return W_{GM}

Since (FL) is a collaborative form of machine learning where the training data is distributed over the clients rather than being stored in a central location. Therefore, the (ViT) model will be used by these clients to train the model with the assigned dataset they have. After training the local model, the average weights are shared with the global model. Then, the global model is updated based on those average weights of all local clients' models. The updated global model weights are then transmitted back to all clients. These steps are repeated until the model accuracy is deemed sufficient [19].

3.4 Applying active learning

We applied an (AL) approach to our model to investigate its impact and enhance the accuracy of the model. Algorithm 1 depicts Active Learning in federated environments for the vision transformer Model. For this, we used pool-based Active Learning algorithms that extract data from a substantial population of unlabeled data. Uncertainty sampling was applied by employing the least confidence sampling which returns the uncertainty score of an array using the least confidence sampling in a 0–1 range where 1 is the most uncertain [14].

Uncertain sampling is one of the most prominent and commonly utilized Active Learning approaches. Dynamic learning utilizes uncertainty sampling to query the most uncertain samples, the instances in which the trainer is least the least confident regarding the classification. It makes decisions based on probabilities and is frequently simple for probability learning approaches [28]. For the Active Learning process in this work, we started by training our local models on an initial labeled pool of 500 images and then performed inference on the unlabeled pool of images and queried images using the least confident sampling. The maximum number of queried images was set at 200 in our case. Those images were labeled, removed from the unlabeled pool of images and then used in the training process once again. We repeated this process for a N_LOCAL_EPOCHS times.

To calculate the Least Confidence Sampling there are two steps:

- Calculate Normalized Uncertainty over N number of classes array. Normalized uncertainty can computed by the following formula in Eq. 1:

$$1 - Max(Predict(NCR)) * (NC/NC - 1) \tag{1}$$

- Get NAL least confident prediction from N number of classes array. We can obtain number of active learning samples (NAL) least confident predictions based on the unlabeled dataset to label it and then add to the model.

  NCR represents the number of classes array, while NC represents the number of classes.

3.5 Ensembling using local and global models

In this paper, we proposed a scheme to integrate the global model with a local model for each client, using an ensemble, which is then used to make inferences on the testing set. We used a weighted average ensemble which refers mainly to soft ensembling. At the level of the local machine, each client trains a vision transformer model on its local dataset. Once these models are trained for each client, their weights (learned parameters) are averaged in order to enable the global model to get the final best weights. Finally, the local model and global model, for each client, are integrated using the ensemble scheme to make final inference on test dataset.

The proposed weighted average ensemble model is shown in Fig. 4 where the following steps are carried out to perform the ensembling:

- The global and local models for each client are applied on the test set to get the inference probabilities as output from each model.
- All probabilities are added and then divided by the number of models used in order to garner probabilities ranging between 0 and 1.
- The class with the highest probability from each prediction is taken.

The complete functioning of the ensemble is described in algorithm 2.

Algorithm 2 Ensemble showing the integration of Active Learning and Federated Learning

Input: transformed_features ϵ;
Define:
ϵ: transformed_features,
α: prediction_of_Active_Learning,
β: prediction_of_Federated_Learning,
A: prediction_of_MajorityVoting
Output: event_type **A**
procedure MULTI_ CLASSIFICATION(ϵ)
$A \longleftarrow$ MajorityVoting(α,β) $\triangleright$ depicts the final prediction of event type
end procedure

4 The evaluation approach

We evaluated the proposed framework for emergency event detection in order to analyze its performance. First, the required environment was set to analyze the performance of the proposed framework.

4.1 Experimental setup

Our models were designed using different Python packages and libraries like Pytorch, numpy, opencv, matplotlib, Scikit-Learn, Keras and native TensorFlow. The hardware used for the model training and testing is Nvidia Tesla P100 (16 GB) offered by Google Colab Pro. Experiments are implemented using the TensorFlow environment on Google Colabotary, with 16GB RAM and 108GB disk. To evaluate the performance of our proposed approach, we used the same hotspot locations from the Snapchat Map using the developed Node.js. We further augmented the dataset by adding more images from Twitter API related to emergency event locations.

4.2 Performance evaluation

During the training phase, the evaluation parameters play an important role to achieve the targeted (ViT) accuracy. Furthermore, gathering appropriate assessment parameters is a critical component to differentiate and develop a perfect model. After the (ViT) model was applied to our dataset, the accuracy of the proposed model was tested through different performance evaluation measures [15] such as:

- Precision (PR): An accurate positive pattern in a positive class is subtracted from the total number of expected positive patterns [24]. Precision can be calculated using Eq. 2.

$$Prec = \frac{TP}{TP + FP} \tag{2}$$

- Recall (RE): The percentage of positively categorized patterns is determined by this measure [24]. Recall can be calculated using Eq. 3.

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

- True Positive (TP): It corresponds to the quantity of cases correctly recognized by the classifier [9].
- False Positive (FP) It represents the amount of negative incidents that were incorrectly classified as positive cases [9].
- False Negative (FN): It relates to the quantity of positive incidents that were incorrectly labelled as negative cases. [9].
- True Negative (TN): The number of negative cases successfully categorised by the model is denoted by the true negative values [9].
- Accuracy (ACC): An accurate classification ratio is calculated by comparing the number of instances analyzed with the number of accurate classifications. Equation 4 can be used to calculate accuracy[9].

$$Accuracy = \frac{(TP + TN)}{TP + FP + TN + FN} \tag{4}$$

- F-Score (FS): Also recognized as the F1-score, this statistic is used to assess data correctness. In addition, it is employed to investigate binary classification methods that

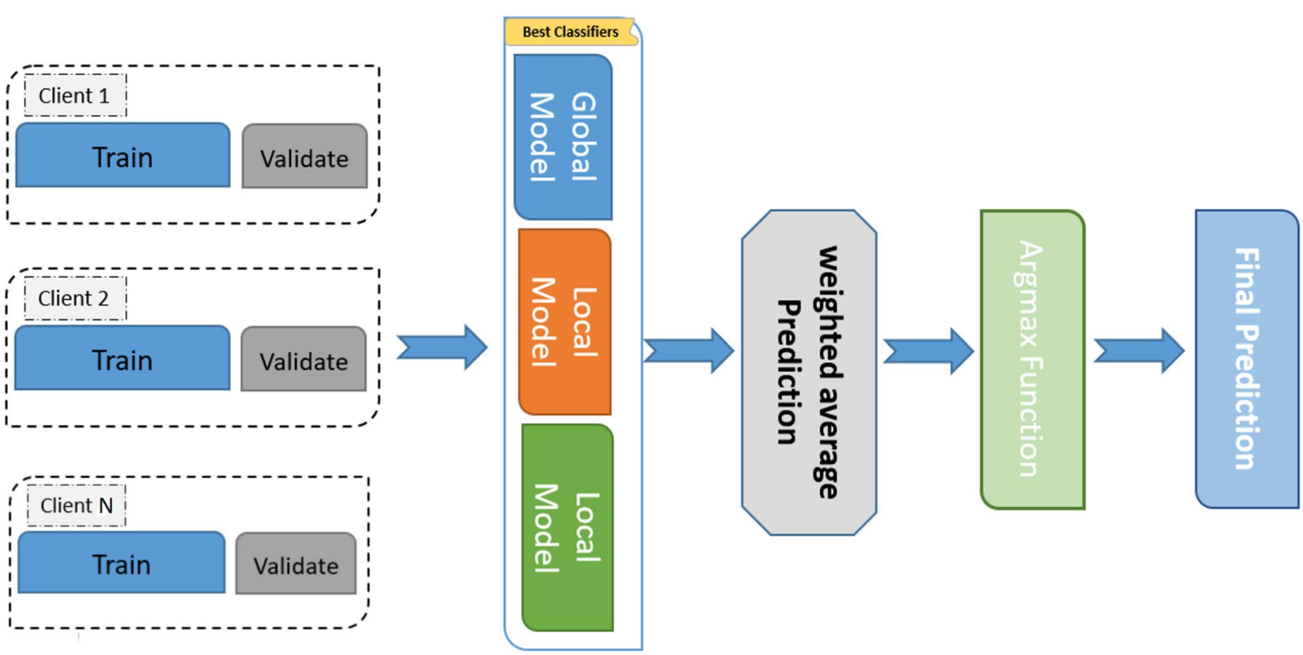

Fig. 4 The proposed weighted average ensemble model

Table 4 Model performance with federated learning without using active learning

Model	Precision	Recall	F-score	Accuracy	Logloss
Before active learning	0.982	0.982	0.982	0.983	0.072

Table 5 Model performance with federated learning and active learning

Model	Precision	Recall	F-score	Accuracy	Logloss
After active learning	0.985	0.985	0.985	0.986	0.062

classify data as "positive" or "negative" [24]. The F1-score can be calculated using Eq. 5.

$$F1 = 2 \times \frac{(precision \times recall)}{precision + recall} \tag{5}$$

- Logloss: It is essential to classify data based on logloss. The following equation illustrates how it can be calculated using probabilities and it is used to compare models. Lower logloss indicates a better-performing model. This value will be sorted in descending order [1].

$$-\frac{1}{N}\sum_{i=1}^{N} y_i \cdot log(p(y_i)) + (1 - y_i) \cdot log(1 - p(y_i)). \tag{6}$$

4.3 Results

The experimental testbed is planned in a way to study the impact of various configurations in our experiment, i.e.

1. The impact of Federated Learning without Active Learning;
2. The impact of Federated Learning with Active Learning;
3. The impact of Ensemble with Federated Learning and without Active Learning;
4. The impact of Ensemble with both Federated and Active Learning.

First, in order to evaluate the proposed approach, we conducted the experiments on the local environment set up described in the previous subsection. Table 4 shows the results of the proposed transformer model performance without applying Active Learning. The results are further enhanced after applying Active Learning. As seen in Table 5, the transformer model achieved a higher accuracy of 98.5% for Precision, Recall and F-score and 0.062 for Logloss.

The performance of each local client is separately measured in our Federated Learning (FL) approach. However, as Figs. 4, 5, 6, 7, 8, 9, 10, 11, 12, 13 and 14 clearly show, the performance of each client approaches 98%. Local client 1 performs well based on the training and validation dataset. As shown in Fig. 5 the accuracy of this client reached 97.4%, while F1 Score was 97.7% as shown in Fig. 6. Figures 7 and 8 depict the precision and recall values. Moreover, Fig. 9 indicates the score for the Logloss evaluation for Local Client 1.

Figures 10 and Fig. 11 present the accuracy and F-score for Client-2 respectively. Both values approach nearly 98%. Similarly, the Figs. 12 and 13 show the precision and recall results for Local Client 2 whereas Fig. 14 indicates the result of Logloss evaluation score for Local Client 2.

In addition, we observe from Table 6 that the proposed ensemble model outperforms the other models in all evaluation metrics as it yields a very high accuracy of 99% for all evaluation metrics.

We further investigated the impact of using Active Learning with the Ensemble method. The results are summarized in Table 7.

5 Conclusion, discussion and future directions

The speed at which information is exchanged has reduced considerably. Furthermore, nowadays, the social media represent the first source of information by providing instantaneous information in real time. Hence, detection of emergency events can provide the entire community with precious information and help people who happen to be close by to take appropriate action, thereby influencing their decision as well as their likelihood of survival. The fundamental benefit of using deep learning to classify and detect problems is that the entire network is trained from raw data to final classification results, which decreases the requirements to build an appropriate feature extraction strategy [18]. Moreover, in social networks, the data labeling procedure is particularly expensive and time-consuming due to a large amount of data available on such platforms, which limits the implementation of deep learning technologies.

In this paper, we employed and analyzed the efficacy of utilizing Active Learning techniques with FL for emergency events using a dataset from social networks to address these limitations. We created our own dataset of images, using the Snapchat API and Twitter, where the images shot at the sites of the emergency events were collected. After collecting images, a FL paradigm was used to split the data between different local clients to train the transformer vision model in order to detect and classify emergency events. Moreover,

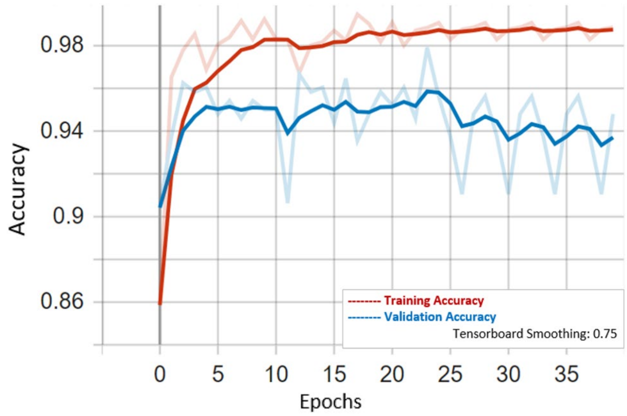

Fig. 5 Accuracy result for Client 1

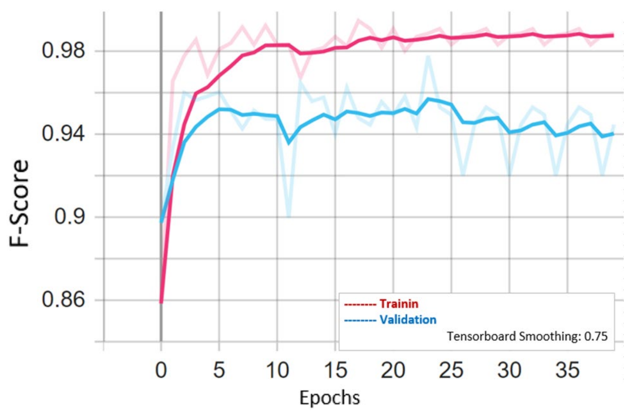

Fig. 6 F-Score result for Client 1

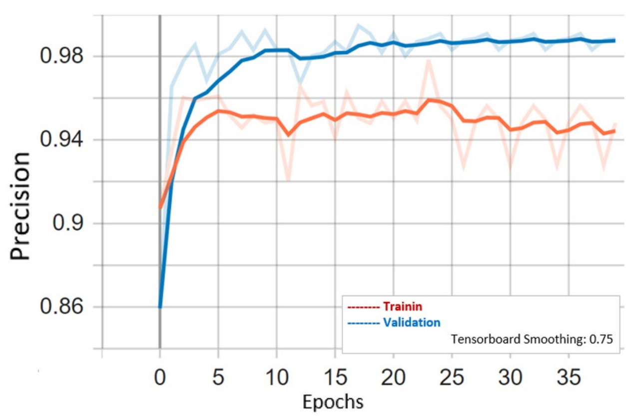

Fig. 7 Precision result for Client 1

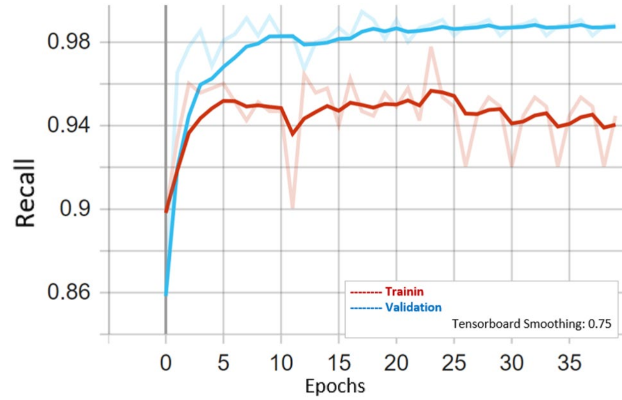

Fig. 8 Recall result for Client 1

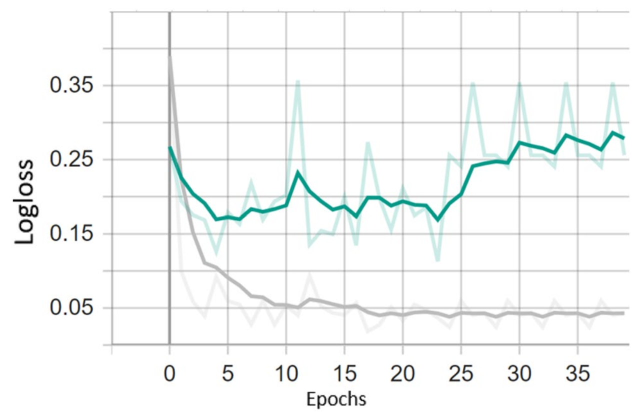

Fig. 9 Logloss result for Client 1

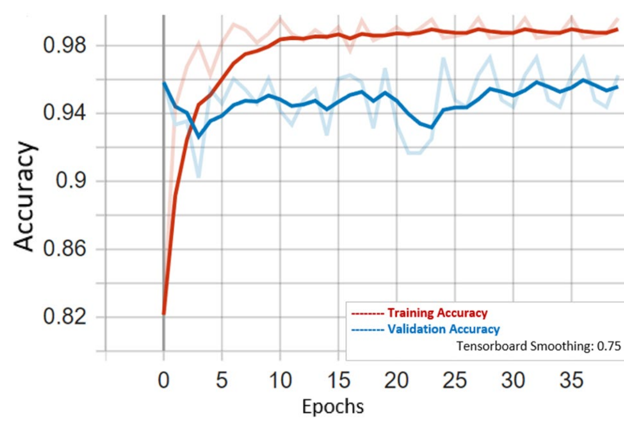

Fig. 10 Accuracy result for Client 2

this paper proposed to use AL methods to select a subset of the whole dataset to accelerate the training process. We further improved the classification results, by employing an ensemble-based technique to integrate the results from the local models (trained on clients' machines) and global model (after aggregating all of the clients' results). The ensemble leaning model yielded a very high accuracy of 99% for all evaluation metrics and the Logloss score reached 0.043.

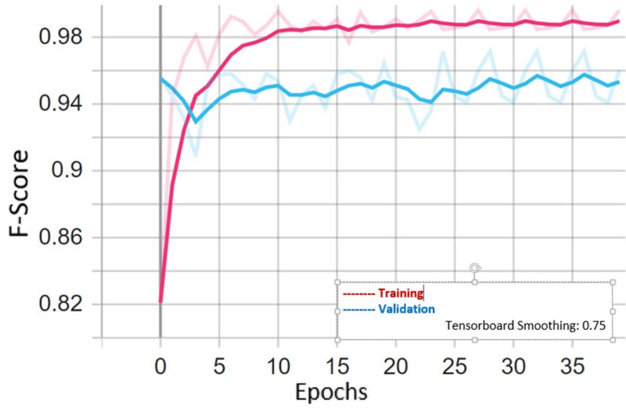

Fig. 11 F-Score result for Client 2

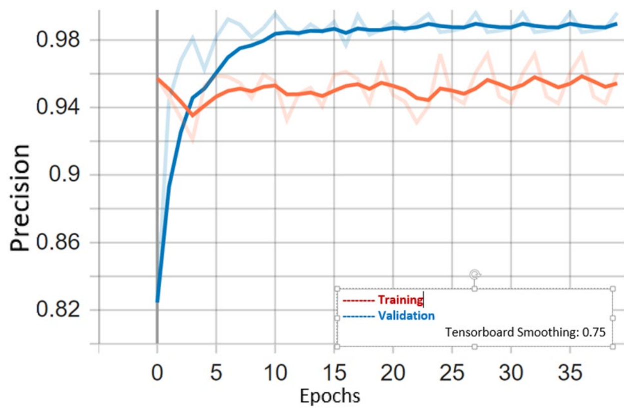

Fig. 12 Precision result for Client 2

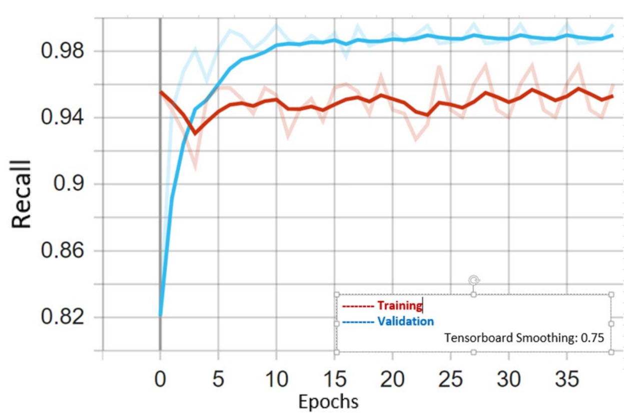

Fig. 13 Recall result for Client 2

These results clearly show the outstanding performance of the proposed approach.

In the future, the proposed technique may be expanded to assess the risk presented by various types of emergency events. It could also be beneficial to add more feature extractions to the approach to improve performance. We also plan to enhance this work by exploring IoT sensors and satellite imagery sources in addition to collecting data from social networks. Further research could focus on optimizing the performance of this approach by exploring different Federated Learning and Active Learning algorithms and expanding the dataset to include a greater diversity of emergency events.

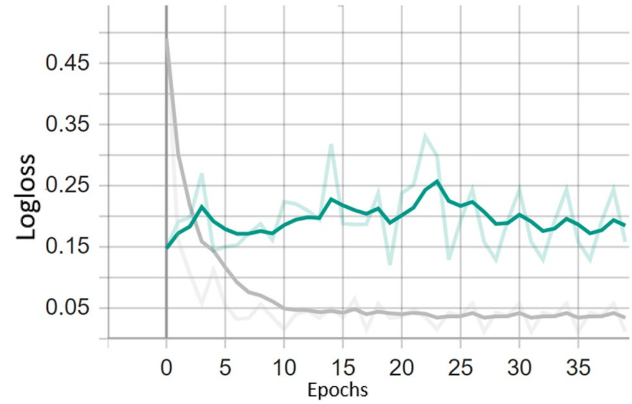

Fig. 14 Logloss result for Client 2

Table 6 Results of model performance after applying the ensemble approach

Model	Precision	Recall	F-score	Accuracy	Logloss
After ensemble learning	0.99	0.99	0.99	0.99	0.043

Table 7 Model performance of applying ensemble with and without active learning

	Accuracy	Logloss	F1-score
With active learning	0.984	0.07	0.984
Without active learning	0.978	0.11	0.978

References

1. Abhinav A, Shiva K, Zekun X, Oluwaseyi F, Nathanael T (2021) Label inference attacks from log-loss scores. In: International conference on machine learning, PMLR, pp 120–129
2. Ahmed L, Ahmad K, Said N, Qolomany B, Qadir J, Al-Fuqaha A (2020) Active learning based federated learning for waste and natural disaster image classification. IEEE Access 8:208518–208531
3. Alfalqi K, Bellaiche M (2022) An emergency event detection ensemble model based on big data. Big Data Cogn Comput 6(2):42

4. Bommel JR (2021) Active learning during federated learning for object detection. B.S. thesis, University of Twente
5. Bügel U, Zielinski A (2013) Multilingual analysis of twitter news in support of mass emergency events. Int J Inf Syst Crisis Response Manag (IJISCRAM) 5(1):77–85
6. Cao J, Yuan X-L, Cui Y, Fan J-C, Chen C-L (2022) A VPN-encrypted traffic identification method based on ensemble learning. Appl Sci 12(13):6434
7. Chaudhary MT, Piracha A (2021) Natural disasters-origins, impacts, management. Encyclopedia 1(4):1101–1131
8. Da Y, Ji Z, Zhou Y (2022) Building damage assessment based on Siamese hierarchical transformer framework. Mathematics 10(11):1898
9. Grandini M, Bagli E, Visani G (2022) Metrics for multi-class classification: an overview. arxiv preprint. 2020: 1–17
10. Guo Y, Wang C, Yu SX, McKenna F, Law KH (2022) Adaln: a vision transformer for multidomain learning and predisaster building information extraction from images. J Comput Civ Eng 36(5):04022024
11. Han K, Wang Y, Chen H, Chen X, Guo J, Liu Z, Tang Y, Xiao A, Xu C, Xu Y et al (2022) A survey on vision transformer. IEEE Trans Pattern Anal Mach Intell
26. Health Services Provider (2020) Emergency and disaster-management. https://healthstandards.org/standard/emergency-and-disaster-management/
12. Horváth J, Baireddy S, Hao H, Montserrat DM, Delp EJ (2021) Manipulation detection in satellite images using vision transformer. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 1032–1041
13. Hossain HMS, Al Khan MA, Hafiz R Nirmalya (2017) Active learning enabled activity recognition. Pervasive Mob Comput 38:312–330
14. Hossin M, Sulaiman MN (2015) A review on evaluation metrics for data classification evaluations. Int J Data Min Knowl Manag Process 5(2):1
15. Kaur N, Lee C-C, Mostafavi A, Mahdavi-Amiri A (2023) Large-scale building damage assessment using a novel hierarchical transformer architecture on satellite images. Comput Aided Civ Infrastruct Eng
16. Khalifa NE, Loey M, Mirjalili S (2021) A comprehensive survey of recent trends in deep learning for digital images augmentation. Artif Intell Rev 55:2351–2377
17. Khanday AMUD, Rabani ST, Khan QR, Rouf N, Din MMU (2020) Machine learning based approaches for detecting COVID-19 using clinical text data. Int J Inf Technol 12:731–739
18. Kim YJ, Hong CS (2019) Blockchain-based node-aware dynamic weighting methods for improving federated learning performance. In: 2019 20th Asia-Pacific network operations and management symposium (APNOMS), IEEE, pp 1–4
19. Li X, Qu Z, Tang B, Lu Z (2021) Stragglers are not disaster: a hybrid federated learning algorithm with delayed gradients. arXiv preprint arXiv:2102.06329
20. Mandavilli VS, Madamala N (2018) Detection of natural disaster affected areas using R. Int J Inf Technol 10:123–131
21. Martínez-Rojas M, Pardo-Ferreira MDC, Rubio-Romero JC (2018) Twitter as a tool for the management and analysis of emergency situations: a systematic literature review. Int J Inf Manag 43:196–208
22. McMahan B, Ramage D (2017) Federated learning: collaborative machine learning without centralized training data. https://www.ai.googleblog.com/2017/04/federated-learning-collaborative.html
23. Miguel J, Caballé S, Xhafa F, Prieto J (2015) A massive data processing approach for effective trustworthiness in online learning groups. Concur Comput Pract Exp 27(8):1988–2003
24. Mittal V, Jahanian M, Ramakrishnan KK (2021) Flare: federated active learning assisted by naming for responding to emergencies. In: 2021 IEEE 29th international conference on network protocols (ICNP), IEEE, pp 1–2
25. Priyanshu A, Sinha M, Mehta S (2021) Continual distributed learning for crisis management. arXiv preprint arXiv:2104.12876
27. Ragini JR, Anand PMR, Bhaskar V (2018) Big data analytics for disaster response and recovery through sentiment analysis. Int J Inf Manag 42:13–24
28. Said N, Ahmad K, Conci N, Al-Fuqaha A (2021) Active learning for event detection in support of disaster analysis applications. Signal Image Video Process 15(6):1081–1088
29. Shastri S, Singh K, Kumar S, Kour P, Mansotra V (2021) Deep-LSTM ensemble framework to forecast covid-19: an insight to the global pandemic. Int J Inf Technol 13:1291–1301
30. Simeon AI, Edim EA, Eteng IE, Ukegbu CC (2021) Design of a flood magnitude prediction model using algorithmic and mathematical approaches. Int J Inf Technol 13(4):1569–1579
31. Simon T, Goldberg A, Adini B (2015) Socializing in emergencies-a review of the use of social media in emergency situations. Int J Inf Manag 35(5):609–619
32. Walambe R, Marathe A, Kotecha K (2021) Multiscale object detection from drone imagery using ensemble transfer learning. Drones 5(3):66
33. Wang L, Yang Y, Li R (2021) Buildformer: automatic building extraction with vision transformer. arXiv preprint arXiv:2111.15637
34. Wang Z, Brenning A (2021) Active-learning approaches for landslide mapping using support vector machines. Remote Sens 13(13):2588
35. Xia Z, Li Z, Bai Y, Yu J, Adriano B (2022) Self-supervised learning for building damage assessment from large-scale XBD satellite imagery benchmark datasets. arXiv preprint arXiv:2205.15688
36. Xu Z, Liu Y, Yen NY, Mei L, Luo X, Wei X, Hu C (2016) Crowdsourcing based description of urban emergency events using social media big data. IEEE Trans Cloud Comput 8(2):387–397
37. Xiang Yu, Wang J, Hong Q-Q, Teku R, Wang S-H, Zhang Y-D (2022) Transfer learning for medical images analyses: a survey. Neurocomputing 489:230–254
38. Zhan X, Wang Q, Huang K-H, Xiong H, Dou D, Chan AB (2022) A comparative survey of deep active learning. arXiv preprint arXiv:2203.13450
39. Zhang C, Yu X, Bai H, Yu B, Li W, Gao Y (2021) A survey on federated learning. Knowl Based Syst 216:106775
40. Zhang Z, He N, Li D, Gao H, Gao T, Zhou C (2022) Federated transfer learning for disaster classification in social computing networks. J Saf Sci Resil 3(1):15–23