FedMix: Mixed Supervised Federated Learning for Medical Image Segmentation*

Report By:
Sarveshkumar Purohit
ai22mtech14006@iith.ac.in

Teaching Assistant:
Paketi Divya
ai21resch01001@iith.ac.in

Guided By:
Prof. C Krishna Mohan

ABSTRACT **Federated Learning (FL) aims to enable multiple clients to jointly train a machine learning model without sharing data. Federated Learning has gained significant attention in the medical imaging community as it can resolve the critical issues of insufficient data and data privacy. However, the existing methods for training an image segmentation model have been based on an unrealistic assumption that the training set for each local client is annotated in a similar fashion and thus follows the same image supervision level. To relax this assumption, in this work, Authors have proposed a label-agnostic unified federated learning framework, named FedMix, for medical image segmentation based on mixed image labels. In FedMix, each client updates the federated model by effectively making use of all available labeled data ranging from strong pixel-level labels and weak bounding box labels to weakest image-level class labels.Based on these local models, They have further proposed an adaptive weight assignment procedure across local clients, where each client learns an aggregation weight during the global model update.**

**This Report is made as part of the course project of CS6450 -Visual computing (offered at IITH during Spring 2023), Code is available at Github

INDEX TERMS Federated learning, Medical Image Segmentation, mixed supervision, medical image segmentation, pseudo labeling, adaptive weight aggregation

I. INTRODUCTION

A. MEDICAL IMAGE SEGMENTATION

The representative task for image content analysis can recognize the category and locate the specific areas.
Machine learning techniques, particularly deep learning, have shown promising results in medical image segmentation. Convolutional neural networks (CNNs) are commonly used because they can learn complex patterns and features directly from the image data. Various architectures, such as U-Net, Fully Convolutional Networks (FCNs), and DeepLab, have been proposed and widely adopted for medical image segmentation tasks.
First, a training dataset with annotated images is required to train the segmentation model. **The annotations can be in the form of pixel-level labels, where each pixel is assigned a class label, or region-based labels, such as bounding boxes or contours.**

B. TYPES OF LABELS IN IMAGE SEGMENTATION TASK

1) Pixel-level labels

Pixel-level labels provide a detailed annotation for each pixel in the image, indicating the class or category it belongs to. This type of labeling is the most granular and precise. However, annotators must manually label each pixel, which can be time-consuming and costly.

2) Box labels

Bounding box labels define the rectangular or square boundaries encapsulating the interest regions. Annotators draw bounding boxes around the structures they want to segment. This type of labeling is more straightforward and less time-consuming than pixel-level or region-based labeling. Still, it provides a rough estimate of the location and extent of the regions.

3) Image-level labels

Image-level data in medical image segmentation refers to a type of label that provides information at the image level rather than at a pixel or region level. While image-level labels may not provide as precise or detailed segmentation information as pixel-level or region-based labels, they still have practical utility.

C. FEDERATED LEARNING

Federated Learning (FL) is a machine learning approach that enables multiple clients or devices to train a shared model collaboratively without sharing their raw data. It addresses privacy concerns by decentralizing the data, thus maintaining data security and confidentiality.

D. SEMI-SUPERVISED LEARNING TECHNIQUE

Semi-Supervised learning is a machine learning algorithm representing the intermediate ground between supervised and unsupervised learning algorithms. It uses a combination of labeled and unlabeled datasets during the training period **where the model is trained on the labeled data first and then used to predict labels for the unlabeled data, which is then incorporated into the training process.**

II. PROBLEM STATEMENT AND MOTIVATION

For a fully-supervised semantic segmentation model, the ideal scenario is, we can collect the pixel-level annotated images as much as possible. this scenario is almost infeasible due to the following three reasons:

1) The strict sharing protocol of sensitive patient information between medical institutions.
2) High pixel-level annotation cost.
3) Expert knowledge usually required for annotating medical images is much more demanding and difficult to obtain.

Clients participating in FL may have different labeling budgets. Therefore, there may be a wide range of inter-client variations in label availability. Weak labels are easier to acquire and thus more broadly available compared to pixel-level ones.

III. CHALLENGES

In practice, there can be various types of weak labels. Effectively utilizing the information from these weakly-labeled data with varying levels of label strengths as well as unlabeled data, especially for clients without pixel-level labeled data, would be highly beneficial for improving the federated model's robustness while preventing training instability.

However, given less informative training data, ensuring that the local model updates from clients without pixel-level labels, e.g., weakly-labeled and unlabeled clients positively contribute to the federated model remains a challenge.

IV. EXISTING RELATED WORK AND ITS LIMITATIONS

1) For the inter-client variations, FL has been combined with domain adaptation [1] [2], to learn a more generalizable federated model.
   Limitations-
   These existing work do not consider the variation in supervision availability which is often observed in clinical practice.
2) Some semi-supervised federated learning methods attempt to utilize the unlabeled data in addition to pixel-level labeled images in training. [3]
   Limitations-
   These existing work do not make any use of the other weakly-labeled images.

V. METHODOLOGY AND IMPLEMENTATION APPROACH

To fully utilize every level of labels available at any client, FedMix addresses the challenge through:

1) Pseudo Label Generation and Selection (Sample Refine). [4]
2) Adaptive Aggregation for Federated Model Update (Aggregate).

Algorithm 1: Pseudocode of FedMix

input : $\overline{D}$: the set of training data
parameter: β, λ: hyperparameters for adaptive aggregation
T: maximum federated training rounds
ϵ: threshold for dynamic sample selection
output : $\theta^T_{\xi 1}$: parameters of F_1
$\theta^T_{\xi 2}$: parameters of F_2
$\theta^0_{\xi 1}$, $\theta^0_{\xi 2} \leftarrow$ initialize()
for $t = 1 : T$ **do**
  $\overline{\mathcal{L}} = \{\}$, $\overline{\theta}_{\xi 1} = \{\}$, $\overline{\theta}_{\xi 2} = \{\}$
  for $i = 1 : |\overline{D}|$ **do**
    $F_1, F_2 \leftarrow$ **Download**($\theta^{t-1}_{\xi 1}$, $\theta^{t-1}_{\xi 2}$)
    $(X, Y) \leftarrow D\ [i]$
    $Y_1, Y_2 \leftarrow F_1(X), F_2(X)$
    $(X, \hat{Y}_1, \hat{Y}_2) \leftarrow$ **Sample&Refine**$(X, Y_1, Y_2, Y, \epsilon)$
    $d_i \leftarrow (X, \hat{Y}_1, \hat{Y}_2)$
    $\Delta\theta^t_{i1}$, $\Delta\theta^t_{i2}$, $\mathcal{L}^t_i \leftarrow$ **Update**$(F_1, F_2; d_i)$
    $\overline{\theta}_{\xi 1}$.add($\Delta\theta^t_{i1}$), $\overline{\theta}_{\xi 2}$.add($\Delta\theta^t_{i2}$), $\overline{\mathcal{L}}$.add($\mathcal{L}^t_i$)
  end
  $\theta^t_{\xi 1}$, $\theta^t_{\xi 2} \leftarrow$ **Aggregate**($\overline{\theta}_{\xi 1}$, $\overline{\theta}_{\xi 2}$, $\overline{\mathcal{L}}$; β, λ)
end
return $\theta^T_{\xi 1}$ and $\theta^T_{\xi 2}$

FIGURE 1: Pseudocode of FedMix

A. TRAINING NOTATIONS

The collection of N clients' training data.
$\bar{D} = [D_1, D_2, ..., D_N]$
Where, for a given client i

For pixel-level labeled data

$$D_i^L = [X, Y_{gt}]$$

unlabeled data

$$D_i^U = [X]$$

image-level class labeled data

$$D_i^{img} = [X, Y_{img}]$$

bounding box-level labeled data

$$D_i^{bbox} = [X, Y_{bbox}]$$

B. PSEUDO LABEL GENERATION AND SELECTION (SAMPLE REFINE).

To utilize every available data and ensure reliable local model updates from clients without pixel-level labels, Authors have designed a novel unified framework using every level of the label to amplify and filter useful signals from pseudo supervision.

- FedMix utilizes consistency regularization - cross-pseudo supervision to generate pseudo labels, which are then dynamically filtered and refined before being used for training.
- The training image X is fed to the F1 and F2 models to generate pseudo labels Y1 and Y2, respectively. Y1 and Y2 are then refined, denoted as $\hat{Y}1$ and $\hat{Y}2$

C. ADAPTIVE AGGREGATION FOR FEDERATED MODEL UPDATE (AGGREGATE)

- Each participating client updates the model locally with its training data Di. Finally, the gradient update from each local client $\nabla\theta_i^t$ is sent to the server to update the federated model's parameters by

$$\theta_\epsilon^t \leftarrow \theta_\epsilon^{t-1} + \sum_{i=1}^{|\bar{D}|} w_i \nabla\theta_i^t$$

Where w_i is the aggregation weight of each client, defined below

- w_i is the aggregation weight of each client, defined below

$$c_i \leftarrow \frac{|\, D_i \,|}{\sum_{i=1}^{i=|\bar{D}|} |\, D_i \,|}, d_i \leftarrow \frac{(\mathcal{L}_i)^\beta}{\sum_{i=1}^{i=|\bar{D}|} (\mathcal{L}_i)^\beta}$$

$$w_i \leftarrow \frac{c_i + \lambda d_i}{\sum_{i=1}^{i=|\bar{D}|} c_i + \lambda d_i}$$

VI. EXPERIMENTS AND RESULTS

For Experiments and Results, the Authors have used datasets from the breast ultrasound task, Skin tumor segmentation. Because of resource constraints, I used the Skin tumor segmentation task dataset. Results are reproduced for the HAM10K dataset with epoch value 60 where original epoch value was 300.

TABLE II: Statistics of the HAM10K dataset

Site	Source	# Patients	# Images
Rosendahl	rosendahl	1552	2259
Vidir	modern	1695	3363
	old	278	439
	molemax	3954	3954

FIGURE 2: Statistics for HAM10K Dataset

Images from Rosendahl, Vidirodern, Vidir-old, and Vidir-molemax are represented by client1, client2, client3, and client4, respectively, and C3 owns the least amount of data, is selected as the client with pixel-level labels. The levels of labels on C1, C2, and C4 have been adjusted accordingly under different settings.

Original Results for 300 Epochs:-

Supervision	client1	client2	client3	client4
U,U,L,B	77.9±2.2	80.0±2.8	96.5±0.8	91.5±0.7

Reproduced results for 60 Epochs:-

Supervision	client1	client2	client3	client4
U,U,L,B	77.84	82.40	92.15	89.87

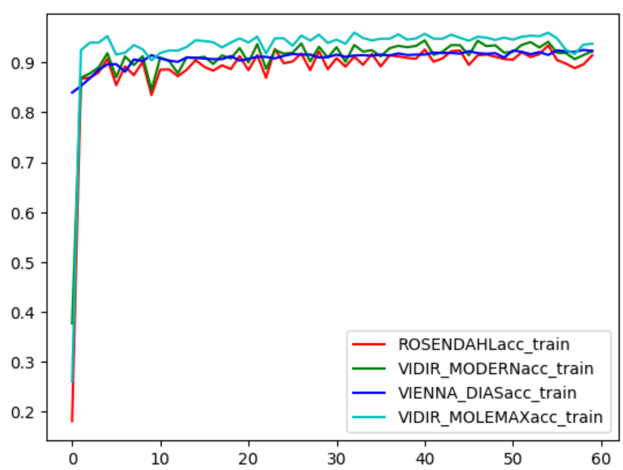

FIGURE 3: Train accuracy vs Epoch

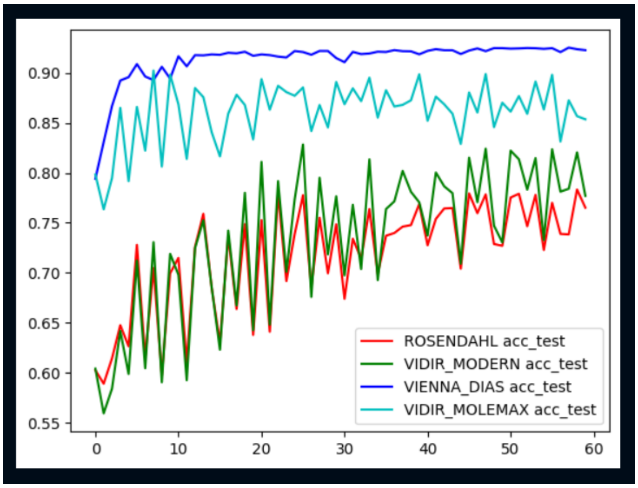

FIGURE 4: Test accuracy vs Epoch

VII. NOVELTY

A. PROBLEM STATEMENT

Suppose there are N clients, denoted c_1, ..., c_N . Each clients has local unlabelled dataset D_i.Our goal is to learn deep learning model for each client with the help of central server, where central server has labelled dataset.So, our aim is to utilize the labelled information available at central server side to train local client having unlabelled dataset with the help of model-constrastive loss.

B. NETWORK ARCHITECTURE

The deep learning network has two components: a base encoder, an output decoder. Each client and central server has same architecture, specifically we are using UNET-3D as deep learning architecture. For ease of presentation, with model weight w, we use $R_w()$ to denote the network before the output layer (i.e., $R_w(x)$ is the mapped representation vector of input x till the encoder part).

C. IMPLEMENTATION DETAILS

At communication round t, suppose client c_i is conducting the local training on unlabelled dataset d_i. It receives the global model w^t from server and updates his model to w_i^t in the local train phase. For every input x, we extract the representation of x from the global model w^t i.e., serv = $R_{w^t}(x)$, representation of x from the local client model of last round w_i^{t-1} i.e., past = $R_{w_i^{t-1}}(x)$, and the representation of x from the local model being updated w_i^t i.e., present = $R_{w_i^t}(x)$. Since the global model should be able to learn better representation with the help of labelled data, our goal is to decrease the distance between present and serv, and increase the distance between present and past.

so we define the model-constrastive loss as,

$$\mathcal{L}_{con} = - \log \frac{\exp\left(sim(present, serv)/\tau\right)}{\exp\left(sim(present, serv)/\tau\right) + \exp\left(sim(present, past)/\tau\right)}$$

where τ is temperature parameter.

same objective can be written as

$$Loss = \mathcal{L}_{con}(w_i^t; w_i^{t-1}; w^t; x)$$

Algorithm 1

Input: Communication round T,number of clients N, Number of local epochs E, temperature τ, learning rate η
Hyperparamter: β, λ for FedMIX weighted average
Output: The final model w^T

for $t \leftarrow 1$ to T **do**
  w^t server weight at t time
  for $i \leftarrow 1$ to N **do**
    $w_i^t \leftarrow \mathrm{w}_t$
    for $i \leftarrow 1$ to E **do**
      clientDataLength.append$(length(D_i))$
      for *each batch* $b = \{x\}$ of D_i **do**
        serv = $R_{w^t}(x)$
        output = $R_{w_i^t}(x)$
        past = $R_{w_i^{t-1}}(x)$
        $loss_i$ =
        $\mu *$ constrativeLoss$(past, output, serv)$
        $w_i^t \leftarrow w_i^t - \eta \nabla loss_i$
      end for
    end for
    loss.append$(loss_i^\beta)$
    localWeight.append(w_i^t)
  end for
  w^{t+1} =
  aggregate$(loss, localWeight, clientDataLength, \lambda)$
end for

The overall federated learning algorithm is shown in algorithm1. In each communication round server sends the global model to the clients,receives the local model from the parties, and updates the global model using weighted averaging fedMIX. In local training each client uses stochastic gradient descent to update the weight which later be send to global model.

After the training of each local client, we are using fedMIX weighted averaging for calculating the final weight w^{t+1} which will be later send to the server.

FedMix presents a novel adaptive aggregation as explained in sectionV-C, unlike FedAVG, FedMix also consider the loss of each client with the dataset length for calculating the weightage given to each localclient weight. FedMix aims to alleviate the training instability which may arise from naively aggregating local client model updates. The weight of each client is determined according to its data quantity and quality (inferred from training loss). In this way, more reliable clients will be assigned with higher weights, leading to better convergence.

Integration of FedMIX weighted averaging is explained in algorithm2

Algorithm 2 aggregate$(loss, localWeight, clientDataLength, \lambda)$

denominatorDI = $sum(loss)$
d = [ ]
for $\mathcal{L}_i$ in $Loss$ **do**
    d.append$(\mathcal{L}_i/denominatorDI)$
end for
denominatorCI = $sum(clientDataLength)$
c = [ ]
for d_i in $clientDataLength$ **do**
    c.append$(d_i/denominatorCI)$
end for
fedmix = [ ]
for $i \leftarrow 1$ `to` N **do**
    fedmix.append$(c[i] + \lambda * d[i])$
end for
denominatorWI = $sum(fedmix)$
w = [ ]
for n_i in $fedmix$ **do**
    w.append$(n_i/denominatorWI)$
end for
for $i \leftarrow 1$ `to` N **do**
    $localWeight[i] = localWeight[i] * w[i]$
end for
w^{t+1} = $sum(localWeight)$
return w^{t+1}

for the calculation of w_i for particular local client, we will first calculate the c_i which will consider the dataset length of local client which is under consideration, then we will calculate d_i, which will consider the training loss of localclient client which is under consideration, and then finally we will calculate the w_i based on d_i and c_i.

λ and β are hyper-parameters to tune, impacting the degree of reliance on different clients. **Thus, adaptive weight assignment according to training loss not only prioritizes learning from more informative clients.**

REFERENCES

[1] "B. n. narayanan and r. c. hardie, "a computationally efficient u-net architecture for lung segmentation in chest radiographs," 2019 ieee national aerospace and electronics conference (naecon), dayton, oh, usa, 2019, pp. 279-284, doi: 10.1109/naecon46414.2019.9058086."

[2] "Z. yan, j. wicaksana, z. wang, x. yang and k. -t. cheng, "variation-aware federated learning with multi-source decentralized medical image data," in ieee journal of biomedical and health informatics, vol. 25, no. 7, pp. 2615-2628, july 2021, doi: 10.1109/jbhi.2020.3040015."

[3] "Q. liu, c. chen, j. qin, q. dou, and p. -a. heng, "feddg: Federated domain generalization on medical image segmentation via episodic learning in continuous frequency space," in proc. cvpr, 2021, pp. 1013-1023."

[4] "X. chen, y. yuan, g. zeng, and j. wang, "semi-supervised semantic segmentation with cross pseudo supervision," in proc. cvpr, 2021, pp. 2613-2622."