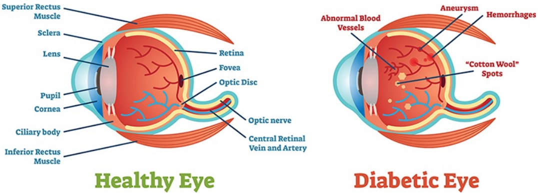
Federated Learning applied to Diabetic Retinopathy prediction

Diego Angulo Ramirez

June 24, 2024

Introduction

Diabetic retinopathy(DR) is a serious complication of diabetes that affects the eyes, potentially leading to blindness if not treated promptly.

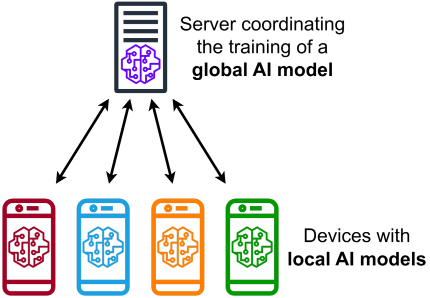


Early detection is of high importance, and with help of Convolutional Neural Networks it is possible to detect the sickness with high accuracy.

Introduction - Why Federated Learning?

ISSUE: Privacy of the data

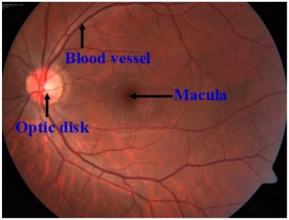
Federated Learning allows us to train a central or global model in a distributed setting, without need of sharing the data in any way.



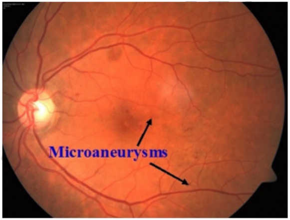
Diabetic Retinopathy

Severity of DR	Lesions observable in fundus image
No DR	No lesions observable
Mild DR	Microaneurysms
Moderate DR	More lesions visible than just microaneurysms.
Severe DR	Any of the following: - Intraretinal bleeding in each of 4 quadrants - Abnormal thickening in the veins of the retina - Prominent intraretinal microvascular abnormalities - No signs of proliferative DR
Proliferative DR	Vitreous/pre-retinal haemorrhages or new vessels.

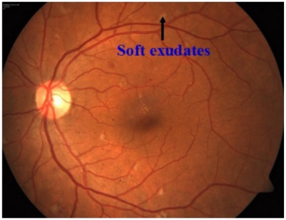
Diabetic Retinopathy



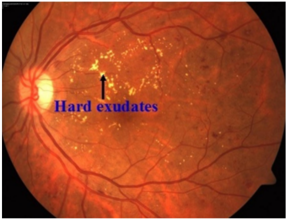
(a)



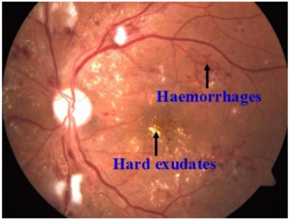
(b)



(c)



(d)



(e)

FedAvg - notation

n	The total number of data points across all clients.
K	Amount of clients, each client has a fixed local dataset, being $\mathcal{P}_k$ the set of indexes of the correspondent data points for client k, also $n_k = \lvert\mathcal{P}_k\rvert$.
C	Fraction of clients selected at the beginning of each round.
E	Number of epochs each client trains over its local data.
B	Size of the minibatches used for training on each client. $B = \infty$ means we use the whole local dataset.
θ, θ^k	Parameters of global model and clients models.

FedAvg - Update of parameters

Usual optimization task for a neural network:

$$\min_{\theta} F(\theta) \triangleq \frac{1}{n} \sum_{i=1}^{n} L(f(x_i;\theta), y_i) \tag{1}$$

Adapted for Federated Setting:

$$\min_{\theta} F(\theta) \triangleq \sum_{k=1}^{K} \frac{n_k}{n} F_k(\theta) \text{ where } F_k(\theta) = \frac{1}{n_k} \sum_{i \in \mathcal{P}_k} L(f(x_i;\theta), y_i) \tag{2}$$

Optimization step(Aggregation):

$$\forall k, \theta_{t+1}^k \leftarrow \theta_t - \eta \nabla F_k(\theta_t) \text{ and for global model } \theta_{t+1} \leftarrow \sum_{k=1}^{K} \frac{n_k}{n} \theta_{t+1}^k \tag{3}$$

FedAvg

Algorithm 1 *FederatedAveraging*[13].

Server executes:

initialize θ_0
for each round $t = 1, 2, \ldots$ **do**
  $m \leftarrow \max(C \cdot K, 1)$
  $S_t \leftarrow$ (random set of m indexes out of the K available clients)
  for each client $k \in S_t$ **in parallel do**
    $\theta_{t+1}^k \leftarrow \text{ClientUpdate}(k, \theta_t)$
  end for
  $m_t \leftarrow \sum_{k \in S_t} n_k$
  $\theta_{t+1} \leftarrow \sum_{k \in S_t} \frac{n_k}{m_t} \theta_{t+1}^k$
end for

ClientUpdate(k, θ) :

$\mathcal{B} \leftarrow$ (split $\mathcal{P}_k$ into batches of size B)
for each local epoch i from 1 to E **do**
  for batch $b \in \mathcal{B}$ **do**
    $\theta \leftarrow \theta - \eta \nabla F(\theta; b)$
  end for
end for
return θ to server

SCAFFOLD - notation

K	Amount of clients.
C	Fraction of clients selected at the beginning of each round.
E	Number of epochs each client trains over its local data.
c^t	Control variate of server after round t.
c_k^t	Control variate of client k after round t.
η_g, η_l	Global learning rate and local learning rate, correspondently.
θ, θ^k	Parameters of global model and clients models.

SCAFFOLD - Update of parameters

For parameters and variates:

$\theta^k \leftarrow \theta^k - \eta_l(\nabla F_k(\theta^k) + c - c_k)$ and $c_k^+ \leftarrow c_k - c + \frac{1}{E\eta_l}(\theta - \theta^k)$

Aggregation:

$$\theta \leftarrow \theta + \frac{\eta_g}{|S|}\sum_{i\in S}(\theta^k - \theta) \quad \text{and} \quad c \leftarrow c + \frac{1}{K}\sum_{i\in S}(c_k^+ - c_k)$$

Main idea behind variates:

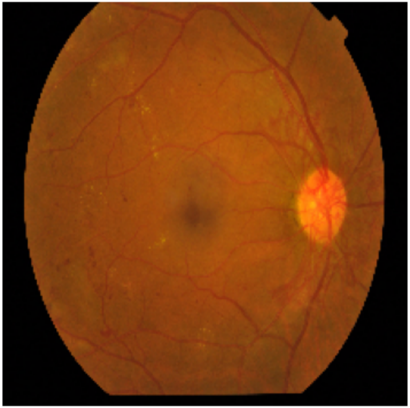
An unbiased local update(no concern about communication costs) would be: $\theta^k \leftarrow \theta^k + \frac{1}{N}\sum_{i=1}^K \nabla F_i(\theta^k)$, as $c_i \approx \nabla F_i(\theta^k)$, $c \approx \frac{1}{N}\sum_{i=1}^K \nabla F_i(\theta^k)$ we can approximate an unbiased update using variates.

SCAFFOLD

Algorithm 2 *SCAFFOLD*[14].

initialize θ, c, and c_k for all clients.
server input: global step-size η_g
client k**'s input:** local step-size η_l
for each round $t = 1, \dots$ **do**
    $m \leftarrow \max(C \cdot K, 1)$
    $S \leftarrow$ (random set of m indexes out of the K available clients)
    communicate (θ, c) to all clients $k \in S$
    for each client $k \in S$ **in parallel do**
        initialize local model $\theta^k \leftarrow \theta$
        for $e = 1, \dots, E$ **do**
            compute mini-batch gradient $\nabla F_k(\theta^k)$
            $\theta^k \leftarrow \theta^k - \eta_l \left(\nabla F_k(\theta^k) - c_k + c\right)$
        end for
        $c_k^+ \leftarrow c_k - c + \frac{1}{K\eta_l}(\theta - \theta^k)$
        communicate $(\Delta\theta^k, \Delta c_k) \leftarrow (\theta^k - \theta, c_k^+ - c_k)$
        $c_k \leftarrow c_k^+$
    end for
    $(\Delta\theta, \Delta c) \leftarrow \frac{1}{|S|} \sum_{k \in S} (\Delta\theta^k, \Delta c_k)$
    $\theta \leftarrow \theta + \eta_g \Delta\theta$ and $c \leftarrow c + \frac{|S|}{K} \Delta c$
end for

Image Pre-processing



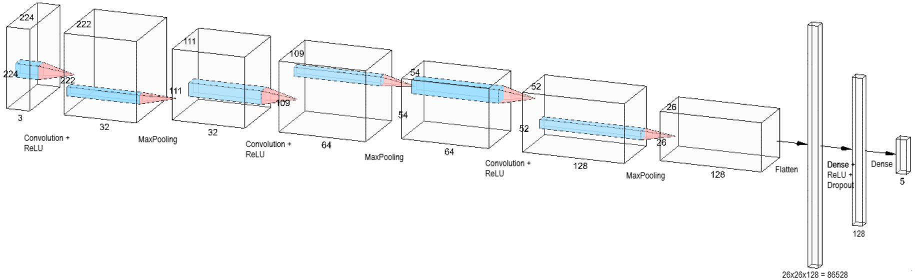
Experiments

APTOS-2019 Dataset was used for the experiments.

4 types of setting were explored:

- Regular CNN + FedAvg
- Regular CNN + SCAFFOLD
- ResNet CNN + FedAvg
- ResNet CNN + SCAFFOLD

Where Regular CNN was the following:



Experiments

Experiment 1: Comparison of the federated models vs all clients

Experiment 2: Best value for C

Results experiment 1 - Regular CNN

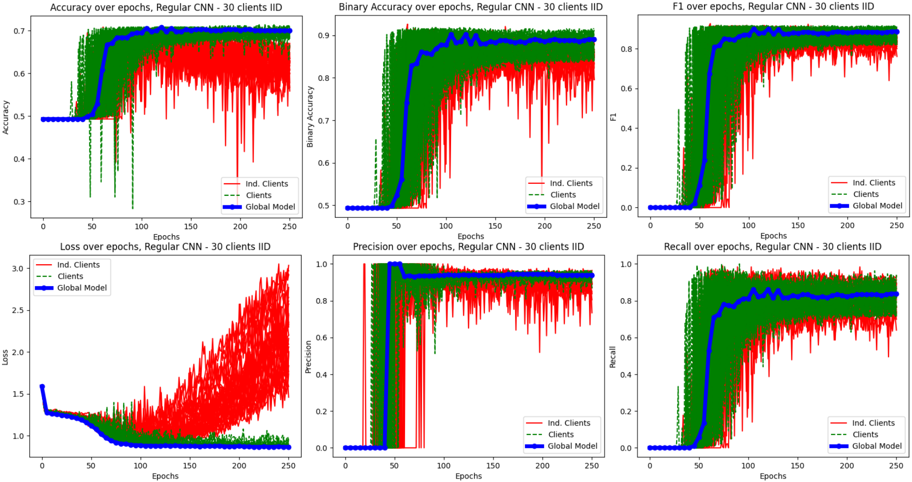


Figure: Performance of Regular CNN + FedAvg on IID data

Results experiment 1 - Regular CNN

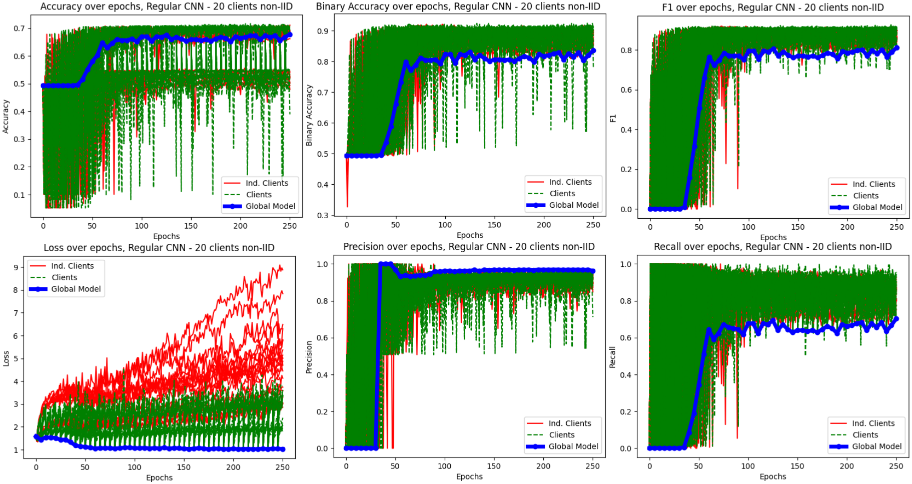


Figure: Performance of Regular CNN + FedAvg on non-IID data

Results experiment 1 - Regular CNN

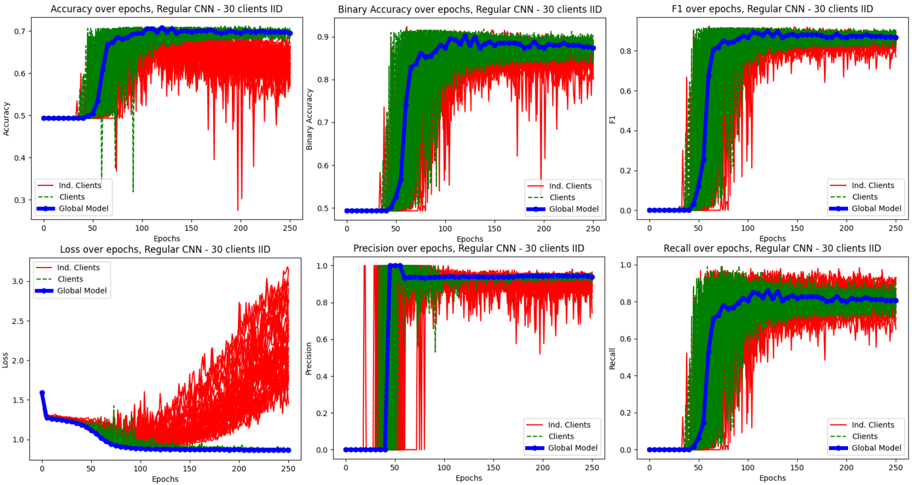


Figure: Performance of Regular CNN + SCAFFOLD on IID data

Results experiment 1 - Regular CNN

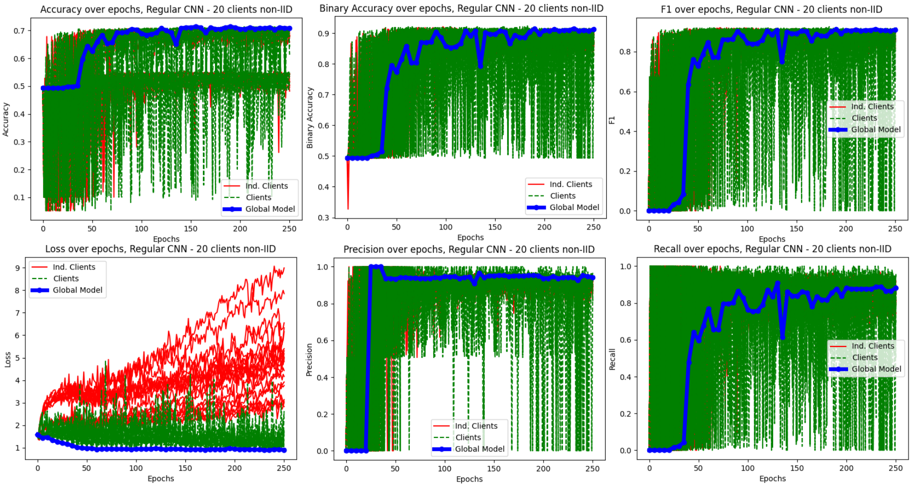


Figure: Performance of Regular CNN + SCAFFOLD on non-IID data

Results experiment 1 - Resnet CNN

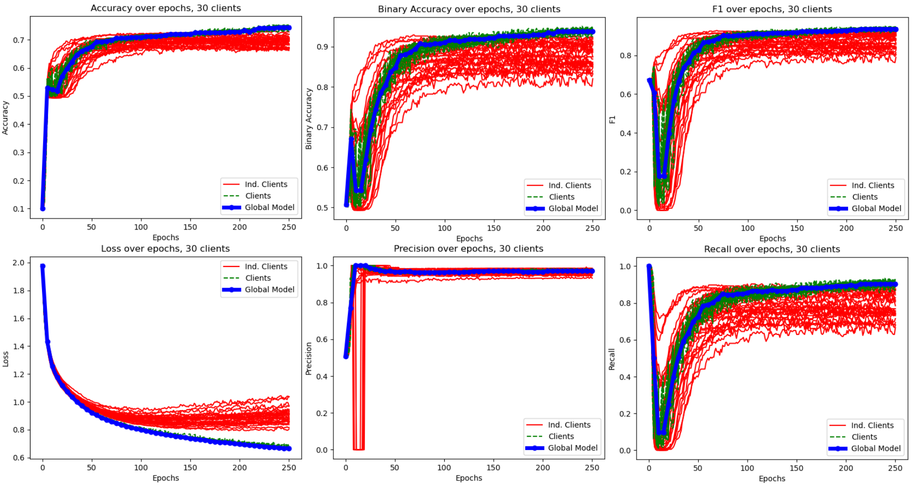


Figure: Performance of ResNet CNN + FedAvg on IID data

Results experiment 1 - Resnet CNN

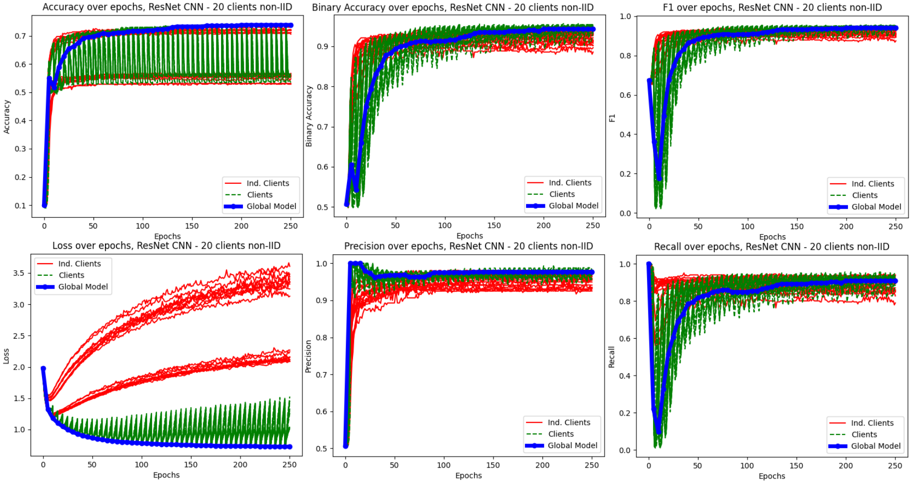


Figure: Performance of ResNet CNN + FedAvg on non-IID data

Results experiment 1 - Resnet CNN

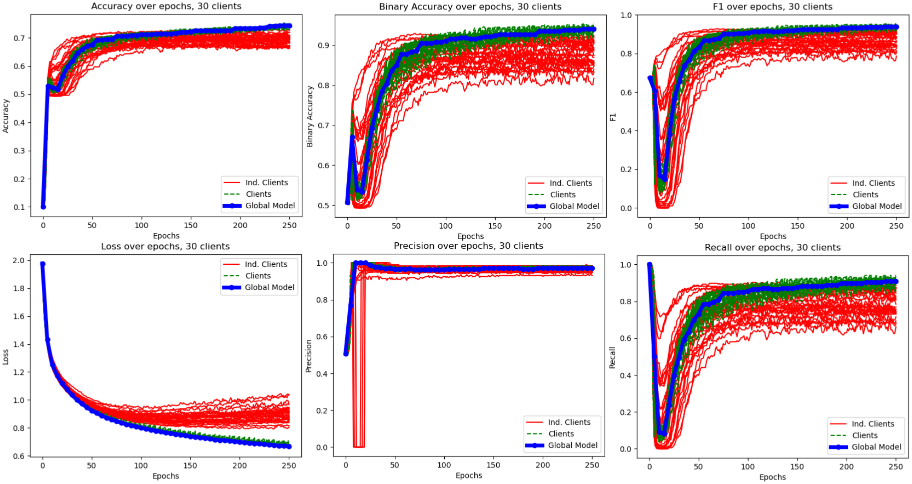


Figure: Performance of ResNet CNN + SCAFFOLD on IID data

Results experiment 1 - Resnet CNN

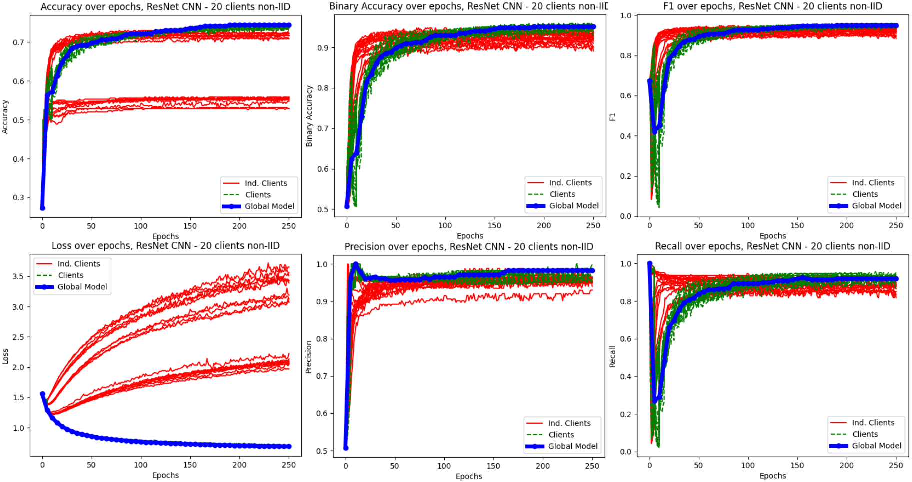


Figure: Performance of ResNet CNN + SCAFFOLD on non-IID data

Results experiment 1

Setting used	Data distr.	Model	Validation data						Test data					
			Loss	Accuracy	Bin. Acc.	F1	Recall	Precision	Loss	Accuracy	Bin. Acc.	F1	Recall	Precision
Regular CNN + FedAvg	IID	Federated	0.8648	70.03%	89.1%	88.64%	83.87%	93.98%	0.8201	70.36%	91.82%	91.71%	89.25%	94.32%
		Independent	1.4586	66.21%	84.74%	83.33%	75.27%	93.33%	1.4052	67.27%	87.64%	87.02%	81.72%	93.06%
	non-IID	Federated	1.0287	67.85%	83.65%	81.37%	70.43%	96.32%	0.9455	69.45%	88.36%	87.3%	78.85%	97.78%
		Independent	2.8671	68.12%	89.37%	89.54%	89.78%	89.3%	2.8771	68.18%	90.73%	91.16%	94.27%	88.26%
Regular CNN + SCAFFOLD	IID	Federated	0.8656	69.48%	87.47%	86.71%	80.65%	93.75%	0.8228	70.18%	91.64%	91.48%	88.53%	94.64%
		Independent	1.4505	66.21%	85.29%	84.3%	77.96%	91.77%	1.4105	67.64%	88.18%	87.76%	83.51%	92.46%
	non-IID	Federated	0.9038	70.84%	91.28%	91.11%	88.17%	94.25%	0.8759	68.36%	90%	90.02%	88.89%	91.18%
		Independent	2.8297	68.12%	89.37%	89.54%	89.78%	89.3%	2.8771	68.18%	90.73%	91.16%	94.27%	88.26%
ResNet CNN + FedAvg	IID	Federated	0.6655	74.39%	93.73%	93.59%	90.32%	97.11%	0.6478	76.18%	96.55%	96.5%	93.91%	99.24%
		Independent	0.8037	68.12%	87.47%	86.39%	78.49%	96.05%	0.7778	69.82%	92.91%	92.57%	87.1%	98.78%
	non-IID	Federated	0.7263	73.84%	94.28%	94.15%	90.86%	97.69%	0.6799	74.36%	96.36%	96.31%	93.55%	99.24%
		Independent	2.0837	73.3%	94.28%	94.18%	91.4%	97.14%	2.1087	73.45%	94.22%	94.21%	90.91%	97.76%
ResNet CNN + SCAFFOLD	IID	Federated	0.6681	74.39%	94.01%	93.89%	90.86%	97.13%	0.65	75.64%	96.55%	96.5%	93.91%	99.24%
		Independent	0.8039	68.12%	87.47%	86.39%	78.49%	96.05%	0.773	70%	92.91%	92.57%	87.1%	98.78%
	non-IID	Federated	0.69	74.39%	95.1%	95%	91.94%	98.28%	0.6518	74.18%	96.36%	96.31%	93.55%	99.24%
		Independent	1.9713	73.57%	94.01%	93.82%	89.78%	98.24%	1.957	74.18%	96%	95.93%	92.83%	99.23%

Results experiment 1

Fed. learning approach for DR	Bin acc on validation	Bin acc on test
Chetoui and Akhloufi (FedAvg+Vision Transformers)	95%	
Chetoui and Akhloufi (FedAvg + CNN)	94%	
ResNet + FedAvg	94.28%	96.36%
ResNet + SCAFFOLD	95.1%	96.36%

Table: Comparison to previous results

Results experiment 2 - Regular CNN

Value for C	FedAvg		SCAFFOLD	
	IID	non-IID	IID	non-IID
0.0(1 client)	84	71	72	19
0.2	>150	129	>150	66
0.5	>150	>150	>150	>150
1.0	>150	>150	>150	>150

Table: Values of C for all settings of Regular CNN

Results experiment 2 - ResNet CNN

Value for C	FedAvg		SCAFFOLD	
	IID	non-IID	IID	non-IID
0.0(1 client)	29	13	21	17
0.2	40	30	38	20
0.5	40	36	44	20
1.0	43	41	44	20

Table: Values of C for all settings of ResNet CNN

Conclusions and questions

...