Experiments and Results

Two Approaches

1. CNNs train from scratch on Covid Dataset
2. Transfer learning from pre-trained models (publicity available) + Fine Tuning on Covid dataset

Notebook Setting (Offline)

Model	FL or ML	Precision	Recall	F1	Accuracy	Loss
CNN (Scratch)	ML	96.94%	94.37%	95.63%	98.44%	0.0013046801645677075
	FL	89.86%	94.42%	91.38%	96.06%	0.10101205219204228
Diff %	**ML vs. FL**	7.08%	0.05%	4.25%	2.38%	0.0997073720274745725
CNN (Pretrained)	ML	97.42%	93.17%	95.05	98.44%	0.0013672651553023902
	FL	95.80%	99.70%	97.67%	96.01	0.0802767796752353 5
Diff %	**ML vs FL**	1.62%	6.53%	2.62%	2.43%	0.0789095145199329598
Improve % Pre-trained vs Scratch	**ML**	0.48%	1.2%	0.58%	0%	0.0000625849907346827
	FL	5.94%	5.26%	6.29%	0.03%	0.02073527251680693

A. Client-Server Setting (Online)

FL Server

Client Notebooks

Model	FL or ML	Precision	Recall	F1	Accuracy	Loss
CNN (Scratch)	FL	`0.9615384615384616`	`1.0`	`0.9803921568627451`	`0.96875`	`0.03176596760749817`
Diff %	**Notebook vs. FL Client Server**	0.02%	0.056%	0.011%	0.008%	0.07%
CNN (Pretrained)	FL	`0.9422792022792023`	`1.0`	`0.9702819256608092`	`0.9526041666666667`	`0.2634122781455517`
Diff %	**Notebook vs. FL Client Server**	0.015%	0.01%	0.065%	0.08%	0.18%

B. Client Server Setting (Online)

FL Server

Client Server

Model	FL or ML	Precision	Recall	F1	Accuracy	Loss
CNN (Scratch)	FL	78.57%	88%	83.02%	92.86%	0.0002678545154530809
Diff %	**Notebook vs. FL Client Server**	11.29%	6.42%	8.36%	3.2%	0.1007441976765891991
CNN (Pretrained)	FL					
Diff %	**Notebook vs. FL Client Server**					

Experimental Setup

1. K-fold cross validation experiments
2. Hyper-parameter tuning/optimisation (?)

Metrics:

1. Average of the K -folds (both for training and test sets). In addition to the metrics shown in table (above), the following would be also computed
    a. Confusion Matrix (TP, FP, TN, FN)
    b. Loss and convergence plots (on training sets)

07 Apr 2022

Table -1 :

1. Similar settings across all experiments (as shown in the columns)
    a. The FL training process should be same
        i. create fl_client
        ii. fl_client set model weights (copy.deepcopy(server.global weights))
        iii. local gradients that should be pass per each client to server and server maintain or adds on the fly.
        iv. del fi_client
    b. The clients – data loader that should be same
    c. Hyper-parameters should be same (lr=3e-3 ADAM optimizer in FL)

	ML (Notebook)	**FL (Notebook)**	**FL (Client Server)**
DarkNet	train&test set 80 : 20 Loss: `0.0002619297844148605` Accuracy: `1.0` Precision: `1.0` Recall: `1.0` F1: `1.0` MAP: Epoch to convergence: 20 Test: Loss: `0.0013046801645677075` Accuracy: `0.984375` Precision: `0.9694444444444444` Recall: `0.9436507936507935` F1: `0.9562850729517396`	train&test set 80 : 20 Loss: `0.02034601407406929` Accuracy: `0.994140625` Precision: `0.9813988095238096` Recall: `0.9923363095238096` F1: `0.9867741580773125` MAP: Epoch to convergence: 30 Test: Loss: `0.10101205219204228` Accuracy: `0.9605902777777778` Precision: `0.8986202686202686` Recall: `0.9442526455026455` F1: `0.9138210131545551`	train&test set 80 : 20 Loss: `0.011830077102786163` Accuracy: `0.996031746031746` Precision: `1.0` Recall: `0.9949238578680203` F1: `0.9974554707379136` MAP: Epoch to convergence: 35 Test Loss: `0.03176596760749817` Accuracy: `0.96875` Precision: `0.9615384615384616` Recall: `1.0` F1: `0.9803921568627451`
CHEXNET	train&test set : 80 : 20 Loss: Accuracy: Precision: Recall: F1: MAP: Epoch to convergence:	train&test set 80 : 20 Loss: 0.2 Accuracy: 86% at 25th Precision: 0.9 Recall: 1 F1: 0.95 MAP: Epoch to convergence:	train&test set 80 : 20 Loss: Accuracy: Precision: Recall: F1: MAP: Epoch to convergence:

Covnext	train&test set : 80 : 20 Training: Loss: 0.010672322678146884 Accuracy: 0.9993489583333334 Precision: 1.0 Recall: 0.9973958333333334 F1: 0.9986928104575163 MAP: Epoch to convergence: 20 Test: Loss: 0.0013672651553023902 Accuracy: 0.984375 Precision: 0.9742063492063492 Recall: 0.9317460317460317 F1: 0.9505450926139657	train&test set : 80 : 20 Training: Loss: 0.053820274906077735 Accuracy: 0.9867621527777778 Precision: 0.9852757057826399 Recall: 0.9983333333333334 F1: 0.9917514055367164 MAP: Epoch to convergence: 50 Test: Loss: 0.08027677967523535 Accuracy: 0.9609375 Precision: 0.9579552790759688 Recall: 0.9969135802469137 F1: 0.9767397542647266	train&test set 80 : 20 Loss: 0.28558457642793655 Accuracy: 0.9485677083333334 Precision: 0.9425340761089817 Recall: 0.9950728155339805 F1: 0.9680911446957281 MAP: Epoch to convergence: 30 Test Set Loss: 0.2634122781455517 Accuracy: 0.9526041666666667 Precision: 0.9422792022792023 Recall: 1.0 F1: 0.9702819256608092

Communication Cost – FL Client Server (Production settings)

	Model Payload	Response Times	Convergence analysis
DarkNet	Upload size, (KB/MB --) 4564.15 KB Download size (KB/MB --) 25355.24 KB	upload time download time model update time on server model update time on client	
Covnext	Upload size, (KB/MB --) Download size (KB/MB --)	upload time download time model update time on server model update time on client	