Federated semi-supervised learning and transfer learning-based multi-task approach to detect Covid-19 and lung-segmentation from chest radiography

Mahbub Ul Alam*, Rahim Rahmani

mahbub@dsv.su.se

Stockholm University

FIRST INRIA-DFKI EUROPEAN SUMMER SCHOOL ON ARTIFICIAL INTELLIGENCE

JULY 20-23, 2021

Abstract

This study explores two medical decision-making tasks namely COVID-19 detection and lung area segmentation detection using chest radiography images. We also explored different cutting-edge machine learning techniques such as federated learning, semi-supervised learning, transfer learning, and multi-task learning to explore the issue. To analyze the applicability of computationally less capable edge devices in the IoMT (Internet of Medical Things) system, we report the results using Raspberry Pi devices as accuracy, precision, recall, F_{score} for COVID-19 detection, and average dice score for lung segmentation detection tasks. We also publish the results obtained through server-centric simulation for comparison. The results show that Raspberry Pi-centric devices provide better performance in lung segmentation detection and Server-centric experiments provide better results in COVID-19 detection.

Data

COVID-19 Radiography Database [1, 3]

3616 'COVID-19' positive cases along with 10,192 'Normal', 6012 'Lung Opacity' (Non-Covid lung infection), and 1345 'Viral Pneumonia' images.

JSRT (Japanese Society of Radiological Technology) database [4]

154 nodule and 93 non-nodule chest radiography images with additional information such as patient age, gender, diagnosis (malignant or benign), X and Y coordinates of nodule, simple diagram of nodule location, and degree of subtlety in visual detection of nodules.

Methods

Federated Learning

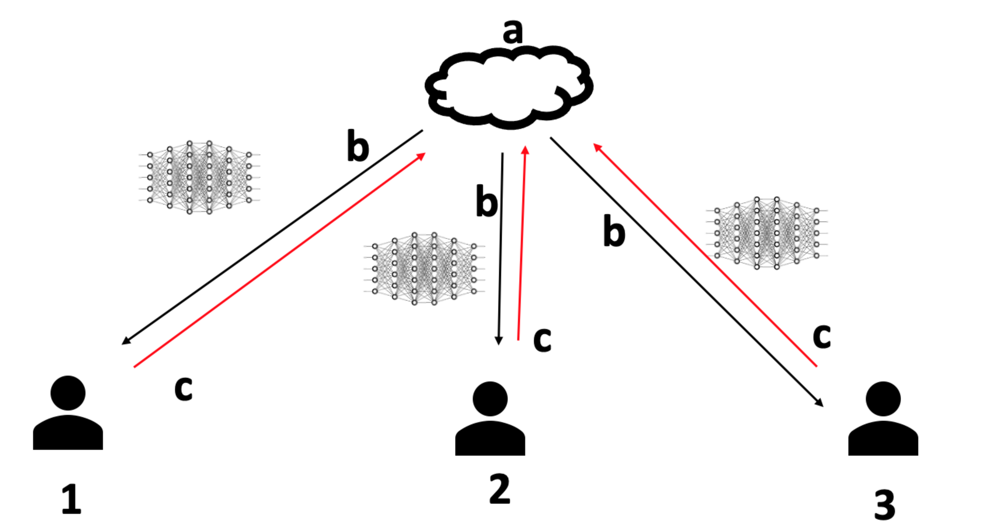

Figure 1: Federated learning architecture: the global model is initialized by the server (a). The server then sends this model to all the client devices (b). After updating the model locally, the clients then send back the updated model to the server (c). The server then aggregates all the local models and updates the global model. One cycle from a-c is known as a round. The process is continued for several rounds.

Simple averaging ('simple')

The global model is calculated as the simple average of all the local model weights.

Standard deviation based weighted averaging ('std_dev')

The local weights to be averaged based on the client's validation metric ('accuracy' or 'loss') on their model with the validation data, and if the metric is greater than the difference between the average of evaluation metrics and standard deviation, then the weights are used for averaging, else the weights are discarded.

Semi-supervised Multi-task learning

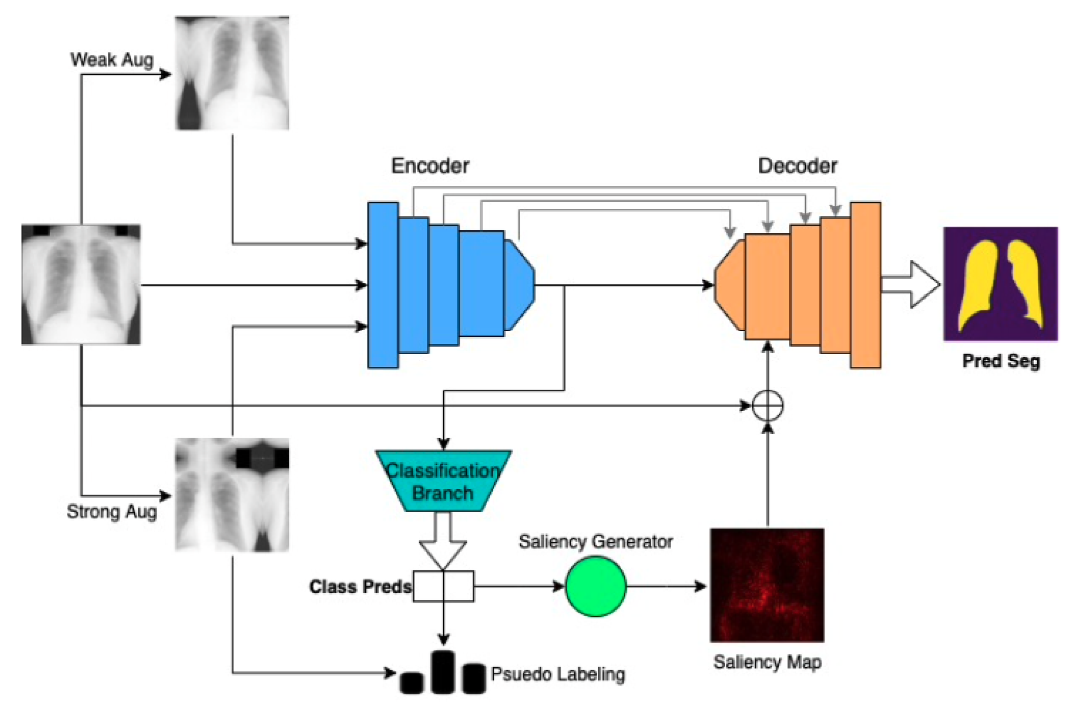

Figure 2: "(classification) Using predictions on unlabeled weakly augmented images, pseudo-labels are generated with confidence, and loss is computed with these labels and the strongly augmented versions of those images. (segmentation) Generated saliency maps from the class predictions are concatenated via the saliency bridge module to guide the decoder for the final segmentations" [2].

Experimental Setup

Figure 3: Eight Raspberry Pi devices were used to conduct the experiments.

- 7632 chest radiography images with equal positive and negative distribution for COVID-19 split into train, test, and validation sets with 80%, 10%, and 10% ratios.
- 246 chest radiography images for segmentation task split into train and test sets with 90% and 10% ratios.
- 50% ratio for number of labeled and unlabeled images.
- Initial learning rate as 0.0001. Adam optimizer having adaptive learning rates of 1.0 was applied in every eight epochs. LReLU with a negative slope of value 0.2. We used 0.25 as the dropout value. The values for t,λ, α, and β were selected as 0.7, 0.25, 5.0, and 0.01.
- Eight different Raspberry Pi 4 devices (Ubuntu 20.10 as GNU/Linux 5.8.0-1024-raspi aarch64) and a A Ubuntu 18.04.5 LTS (GNU/Linux 4.15.0-142-generic x86_64) based server with GeForce RTX 2080 Ti GPU.

Results

Validation Metric	Aggregation Technique	Accuracy	Precision	Recall	F_{score}	Average Dice Score
loss	**std_dev**	**0.720**	**0.666**	**0.883**	**0.759**	**0.844**
loss	simple	0.680	0.637	0.833	0.722	0.769
accuracy	std_dev	0.655	0.677	0.591	0.631	0.708
accuracy	simple	0.717	0.682	0.814	0.742	0.795

Table 1: Results of Server-centric federated semi-supervised learning experiments.

Aggregation Technique	Accuracy	Precision	Recall	F_{score}	Average Dice Score
simple	0.694	0.663	0.789	0.721	0.844
std_dev	0.683	0.633	0.872	0.733	0.785

Table 2: Results of Raspberry Pi-centric federated semi-supervised learning experiments with ten rounds.

Total Epochs	Validation Metric	Accuracy	Precision	Recall	F_{score}	Average Dice Score
10	accuracy	0.647	0.780	0.410	0.538	0.767
10	**loss**	**0.782**	**0.787**	**0.773**	**0.780**	**0.798**
15	**loss**	**0.802**	**0.810**	**0.790**	**0.800**	**0.780**
15	accuracy	0.761	0.817	0.673	0.738	0.708
5	loss	0.697	0.804	0.521	0.632	0.785
5	accuracy	0.625	0.618	0.655	0.636	0.737

Table 3: Results of Server-centric initial model training for transfer learning experiments.

Total Rounds	Valid. Metric	Aggregation Technique	Accuracy	Prec.	Recall	F_{score}	Average Dice Score
5	loss	simple	0.772	0.744	0.829	0.784	0.806
5	accuracy	std_dev	0.809	0.770	0.881	0.822	0.807
5	accuracy	simple	0.730	0.664	0.931	0.775	0.812
5	loss	std_dev	0.749	0.688	0.912	0.784	0.815
10	loss	simple	0.820	0.830	0.804	0.817	0.835
10	**loss**	**std_dev**	**0.827**	**0.788**	**0.895**	**0.838**	**0.828**
10	accuracy	simple	0.759	0.867	0.612	0.717	0.823
10	accuracy	std_dev	0.769	0.875	0.628	0.732	0.837

Table 4: Results of Server-centric federated semi-supervised transfer learning experiments with an initial model obtained through training 15 epochs.

Initial Training Epochs	Aggregation Technique	Accuracy	Precision	Recall	F_{score}	Average Dice Score
10	simple	0.725	0.656	0.949	0.775	0.884
15	simple	0.760	0.697	0.920	0.793	0.855
10	std_dev	0.723	0.651	0.964	0.777	0.883
15	std_dev	0.761	0.699	0.917	0.793	0.854

Table 5: Results of Raspberry Pi-centric federated semi-supervised transfer learning experiments with ten rounds.

Figure 4: Best segmentation dice score (0.941) visualization, left: original image, middle: ground truth, right: predicted segmentation.

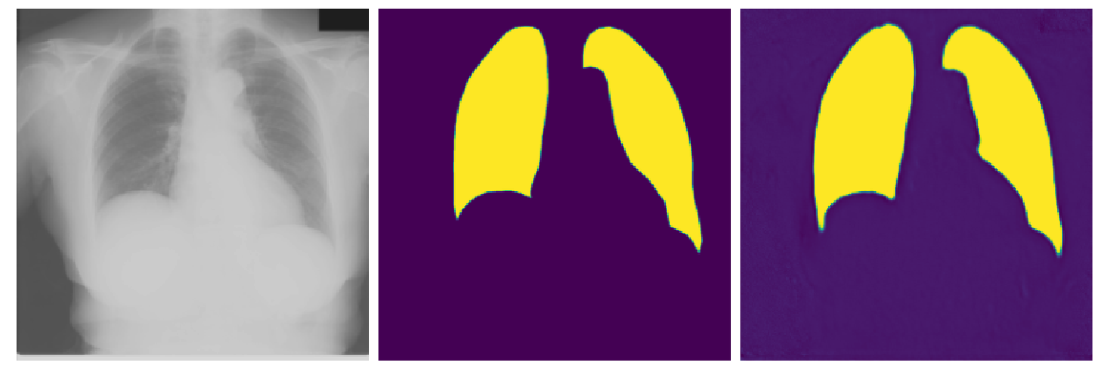

Figure 5: Worst segmentation dice score (0.776) visualization, left: original image, middle: ground truth, right: predicted segmentation.

Conclusions

We investigated the COVID-19 detection and lung segmentation detection problems from chest radiography images using various recent machine learning strategies such as federated learning, semi-supervised learning, transfer learning, and multi-task learning. We have compared the IoMT (Internet of Medical Things) setup deployed using Raspberry Pi devices with high-end computational device-based experiments and found that although for lung segmentation detection Raspberry Pi provides better results, it is slightly worse for COVID-19 detection. We posit that a more nuanced representation of data with the practical construction of a superior IoMT framework would provide a better automatic diagnostic aiding tool for the stakeholders.

Forthcoming Research

Investigation using a multi-modality-based superior data representation combined with real-time data provided by various sensors and superior edge devices.

References

[1] Muhammad EH Chowdhury, Tawsifur Rahman, Amith Khandakar, Rashid Mazhar, Muhammad Abdul Kadir, Zaid Bin Mahbub, Khandakar Reajul Islam, Muhammad Salman Khan, Atif Iqbal, Nasser Al Emadi, et al. Can ai help in screening viral and covid-19 pneumonia? *IEEE Access*, 8:132665–132676, 2020.

[2] Ayaan Haque, Abdullah-Al-Zubaer Imran, Adam Wang, and Demetri Terzopoulos. Multimix: Sparingly-supervised, extreme multitask learning from medical images. In *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, pages 693–696, 2021.

[3] Tawsifur Rahman, Amith Khandakar, Yazan Qiblawey, Anas Tahir, Serkan Kiranyaz, Saad Bin Abul Kashem, Mohammad Tariqul Islam, Somaya Al Maadeed, Susu M Zughaier, Muhammad Salman Khan, et al. Exploring the effect of image enhancement techniques on covid-19 detection using chest x-ray images. *Computers in biology and medicine*, 132:104319, 2021.

[4] Junji Shiraishi, Shigehiko Katsuragawa, Junpei Ikezoe, Tsuneo Matsumoto, Takeshi Kobayashi, Ken-ichi Komatsu, Mitate Matsui, Hiroshi Fujita, Yoshie Kodera, and Kunio Doi. Development of a digital image database for chest radiographs with and without a lung nodule: receiver operating characteristic analysis of radiologists' detection of pulmonary nodules. *American Journal of Roentgenology*, 174(1):71–74, 2000.