Eye Disease Classification using Federated Learning

Shankar Kumar Senthil Kumar*, Ashwin Lamani*
Department of Computer Science
University of Illinois Urbana-Champaign
Champaign, Illinois 61801
sks10@illinois.edu
alamani2@illinois.edu
Code†

A. Introduction and Motivation

Computer vision and deep learning have significantly advanced medical image analysis. Over the years, there has been an increase in the number of datasets for medical images analysis. Using these large collection of medical images, several researchers have been able to achieve state-of-the-art performance in various tasks. Works as [18] [19] have shown the effectiveness of applying deep learning CNN models for the tasks of medical image classification and lesion segmentation. Furthermore, it has been shown that with the help of deep learning, the onset of tumors and other diseases can be identified during the early stage of infection. All these areas have demonstrated significant potential in the use of deep learning and AI in healthcare settings.

Deep convolutional neural networks consist of several layers with millions of parameters. Often these complex and large models require learning from large amounts of data to be able to achieve the desired performance. However, in a real-world medical setting, gathering such large amounts of sensitive data is a herculean task. Further, a model that has only been trained on data from one institute can easily be overfitted, which would lead to poor generalization and a significant bias towards that one institute's data. For instance, a particular skin condition within a specific demographic might exclusively manifest among individuals with fair skin in one institution, whereas in another setting, a different demographic could present instances of the same skin condition among those with darker skin tones. One potential solution to address this data scarcity is to gather imaging data from various sites. This approach may lead to a rise in the quantity and variety of data gathered. However, most medical institutes also have strict rules and policies when it comes to sharing patients' data with external parties.

Emerging as a viable solution for addressing the issue of limited sample sizes and safeguarding personal privacy, federated learning [12] [5] [9] has garnered attention as a prominent research focus in recent times. It aims to collaboratively train machine learning models without the need for sharing or exchanging data across various sites. Federated learning as shown in figure 1 involves multiple clients which train on their local datasets. The local models are gathered by the server, which then computes a global model and broadcasts it to all the clients for deployment.

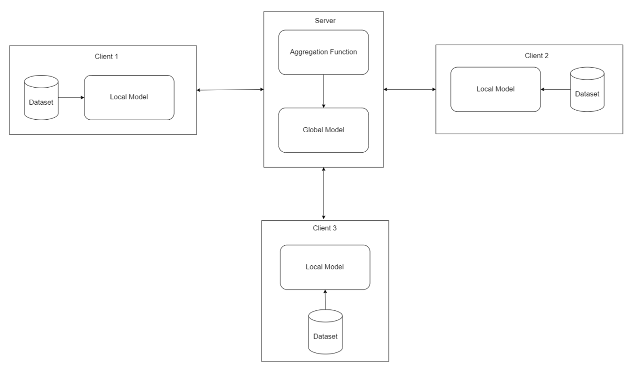

Figure 1. Overview of federated learning. At the start of every round, the server sends the current model parameters to each of the clients. The clients then train their local model on their local dataset and send their trained parameters back to the server. Within the server, the incoming parameters are aggregated to refine the Global model.

Although federated learning has been a promising approach to tackle several problems, the applications of federated learning for medical tasks have been minimal. Recent surveys introduce the applications of federated learning in medicine and healthcare areas [3] [14] [13]. Most survey papers cover broader areas and do not delve into the usage

*These authors contributed equally to this work.
†https://github.com/lashwi/CS444-Project

of federated learning for medical image analysis. Using a CXR dataset from 20 worldwide hospitals and trained FL models, Dayan [10] achieved 95% sensitivity and 88.2% specificity scores across all hospitals to predict clinical outcomes. Lee et al [6] collected 8457 thyroid ultrasound pictures from six universities and trained various deep learning networks (VGG19, ResNet50, ResNext50, SE-ResNet50, and SE-ResNext50) in FL conditions. According to their published findings, under FL circumstances, the area under the receiving operating characteristic (AUROC) stayed between 75.20% and 86.72%, while for standard deep learning models, it ranged from 73.04% to 91.04%.

In this work, we propose a federated learning framework for eye disease classification. We make use of ResNet50 and VGG19 architectures to train a classification model by making use of transfer learning techniques. Further, we compare the performance of these models against simulated federated learning models based on the ResNet50 and VGG19 architectures.

B. Approach

Our proposed federated learning model combines the advantages of large computer vision models with distributed learning to ensure the privacy of data. Federated learning also allows our model to be trained on a diverse set of data obtained from various clients.

B.1. Data Augmentation

Deep learning models are highly complex and require a large number of instances to avoid overfitting. However, high-quality images in a real-world setting are scarce. In the case of medical data, things are even more complicated. Most institutes have regulations protecting patient health data, and as a result, do not share it publicly. In our work, we leverage data augmentation techniques to tackle the issue of data scarcity. The process of data augmentation involved modifying images to increase the number of training examples while preserving semantic information. Our approach uses three transformations on the training data namely: rotation and horizontal flip. These augmentation techniques were carefully chosen to ensure that the visual details of the original image are maintained, and no bias is introduced.

B.2. Data Preprocessing

In deep learning and computer vision applications, preprocessing is an important practice. In our work, we make use of preprocessing to ensure that all retinal images are of the same size to ensure uniformity among the images. The input retinal images are resized to 224x224 to ensure compatibility with the network architectures. Resizing all the input images to a uniform size helps the model converge and generalize faster.

B.3. Standard Multi-Class Classification

The task of classification involves performing multi class classification to distinguish the retinal images into Normal, Diabetic Retinopathy, Cataract or Glaucoma. We make use of two CNN architectures as outlined below for retinal image classification and compare their performance.

B.3.1 ResNet50

ResNet-50 [7] is an expansion of the original ResNet-34 model. The ResNet-50 architecture consists of 48 convolutional layers, one MaxPool layer and one average pool layer. This architecture also consists of a novel method to prevent vanishing gradient by making use of residual connections. In addition, these residual connections enable the model to learn an identity function. This ensures that the higher layers of the model do not perform any worse than the lower layers. ResNets in general have been widely used for various image classification tasks and proven to be highly effective [8] [16]. In our work, the last fully connected layer is removed and three fully connected layers of dimensions (2048,1024), (1024,1024), (1024,4) with ReLu activation are added to the classification model.

B.3.2 VGG19

The VGG19 [17] architecture contains 19 trainable convolutional layers with ReLu activations that are connected in a feed forward fashion. The VGG architecture has been used in several ground-breaking object recognition models and has surpassed baselines on many datasets beyond ImageNet. In our work, the fully connected layers as part of the classification layer are removed and two linear layers of dimensions (4096x1024), (1024x4) with Relu activations between them are added.

B.4. Federated Learning Multi-Class Classification

Federated learning is a decentralized approach to train machine learning models. The advantage of federated learning is that there is no longer centralized training and as a result, data need not be transferred from the client to the server. Instead, the raw data that is present in the clients can be used to train the model locally and then can be aggregated to form a global model. In federated learning, we move the computation to the data rather than moving the data to the computation. Federated Learning has two major steps as outlined below.

B.4.1 Client-Side Training

The initial model is initialized on the server using random initial parameters. These parameters are then sent to each of the connected client nodes. As a result, each of the client nodes now has a copy of the parameters from the server. A local copy of the model is initialized using these parameters and the model is trained on the local dataset. The training can be performed either partially or completely till convergence. Once training is complete, the client returns its updated trained parameters back to the server.

B.4.2 Server-Side Aggregation

The server receives the updated parameters from all of its clients. Each client has a different set of data and as a result, they return different updated parameters. In order to aggregate the various parameters into one single model, an aggregation function is used. In our work, we used Federated Averaging (FedAvg) [11]. FedAvg computes a weighted average of the updated parameters that the server receives. The weights for each of the updated parameters depend on the number of data samples used by the client for training. This ensures that each of the models contribute equally to the aggregated parameters.

The above two steps are performed across various rounds. In the next round, each of the client models is initialized with the newly updated parameters from the server. Since in each round the clients are not fully trained till convergence, the aggregated model will also not be fully trained. Hence, the entire process needs to be repeated across multiple rounds to achieve convergence of the global aggregated model.

C. Experimental Setup

In order to evaluate the efficacy of federated learning, we compared and contrasted the results obtained using pre-trained ResNet-50 and VGG-19 models both in a standard and federated setting. Training and testing of the models were performed using a Nvidia T4 Tensor GPUs along with 32-GB RAM. The proposed deep learning framework has been implemented using Pytorch deep learning library [1] and Flower framework [4] for federated simulations.

C.1. Dataset

We conduct the experiment on the popular benchmark eye classification dataset [2]. This dataset is used since there are multiple classes namely Normal, Diabetic Retinopathy, Cataract and Glaucoma retinal images where each class has approximately 1000 images. There are a total of 4217 images in the dataset across the four classes split uniformly. We split the dataset into training, validation and test set in the ratio 7:2:1 respectively i.e 3036, 759 and 422 images respectively.

C.2. Model Training

For all the experiments, we used a learning rate of 1e-3 with an Adam optimizer. Regularisation techniques such as

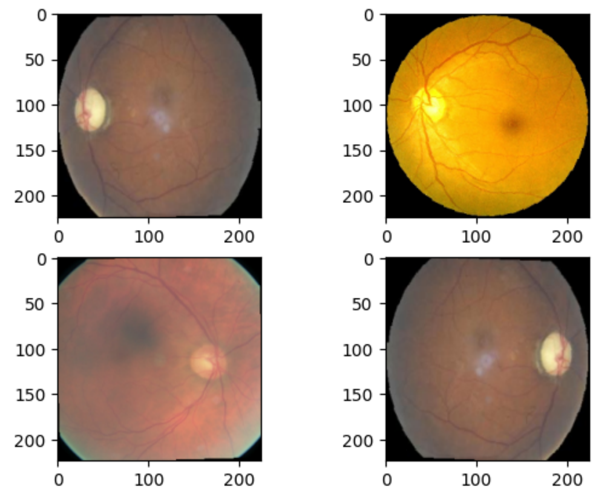

Figure 2. Examples of images from dataset for Normal, Diabetic Retinopathy, Cataract and Glaucoma respectively

early stopping and reducing the learning rate on the plateau were deployed to attain the best performance. Early stopping was used with a patience equal to 5 and it helped prevent overfitting the data. Reducing the learning rate occurred with a patience equal to 3 and this helped the model converge faster. For our experiments, we froze the layers in the following way:

- For partially frozen ResNet-50, we froze all the layers except the last sequential block and the newly added linear layers.
- For fully frozen ResNet-50, we froze all the layers except newly added last linear layers.
- For partially frozen VGG-19, we froze all the convolution layers except the last four convolutional layers and the classifier layer of the model
- For fully frozen VGG-19, we froze all the layers except the newly added linear layers.

C.2.1 Standard Training

We use ResNet-50 and VGG-19 with pre-trained ImageNet weights for our experiments. We leverage the use of transfer learning by freezing the initial layers and only training the latter layers. In our experiments, the partially frozen ResNet-50 model converged on epoch 12, the fully frozen ResNet-50 model converged on epoch 19, the fully frozen VGG-19 model converged on epoch 18, and the partially frozen VGG-19 model converged on epoch 23.

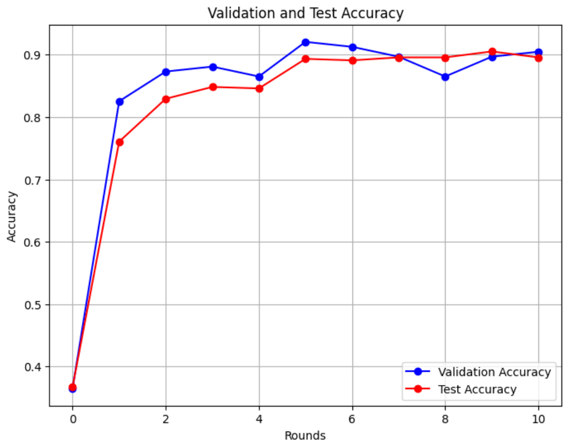

Figure 3. Accuracy Plot for fully frozen ResNet-50 using Federated Learning

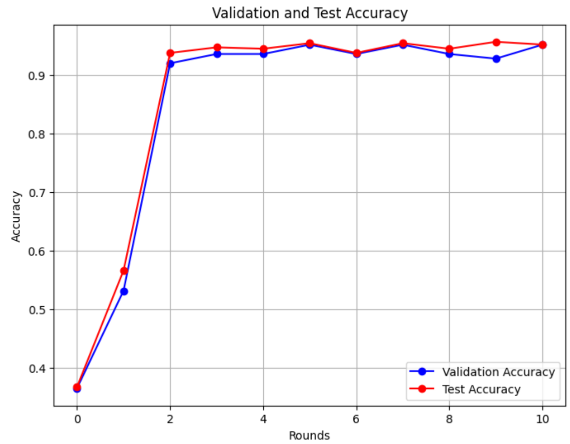

Figure 5. Accuracy Plot for partially frozen ResNet-50 using Federated Learning

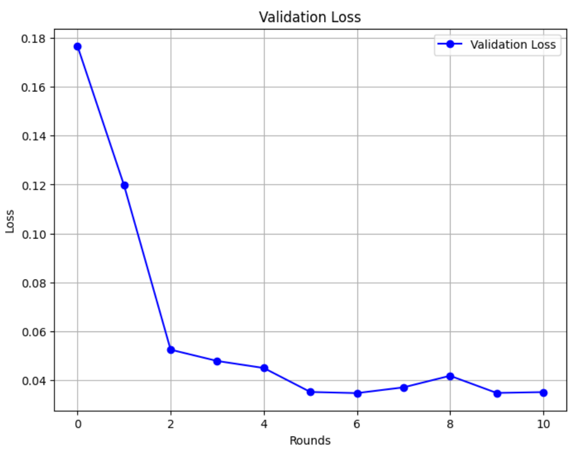

Figure 4. Loss plot for fully frozen ResNet-50 using Federated Learning

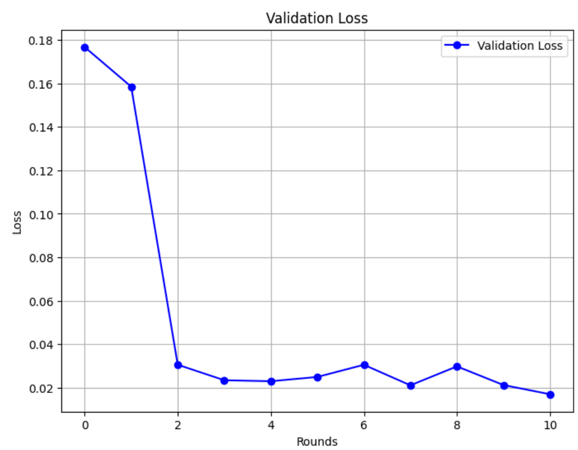

Figure 6. Loss plot for partially frozen ResNet-50 using Federated Learning

C.2.2 Federated Training

To simulate federated learning, we use 3 clients for all our experiments. Other notable parameters used during the experiments for the client are number of rounds equal to 10, the number of CPUs equal to 4, the number of GPUs equal to 2 and batch size equal to 8. The ResNet-50 and VGG-19 models were trained in the federated setting using the best hyperparameters that were obtained from fine-tuning the standard version of the models.

C.3. Preprocessing and Transformation

The 48x48 images in the dataset are first resized to 224x224 across all the images. Post resizing, we augment the dataset with RandomRotation with degree equal to 5 and RandomHorizontalFlip with probability equal to 0.5 to diversify the images.

D. Results

Test Accuracy	Standard	Federated
ResNet fully frozen	88.38	90.52
ResNet partially frozen	**92.18**	**95.73**
VGG fully frozen	88.62	84.59
VGG partially frozen	91.70	26

The table provided offers a comprehensive summary of the outcomes derived from various models under distinct training modes.

Figure 3 shows the accuracy plot for fully frozen ResNet-50 model using federated learning. In comparison Figure 7 shows accuracy for partially frozed ResNet-50 model under standard training. These plots indicate that federated learning can help achieve comparable accuracy to standard training.

It can also be observed from Figure 5 that ResNet-50

Figure 7. Accuracy Plot for partially frozen ResNet-50

model that is partially frozen using federated learning performs better than when fully frozen as seen in Figure 7. This complies with our expectation that the model when partially frozen gets tuned to the image dataset better than when it is fully frozen.

Additionally, we also explored the use of FedAdagrad [15] as an aggregation function. However, we encountered notable challenges characterized by significant training instability. This instability during the training process is likely attributed to the limited volume of available images, which could hinder the model's ability to effectively learn and generalize. The scarcity of data might result in erratic fluctuations or challenges in convergence during the training phase, leading to an unstable learning process within the federated framework.

It is important to note that training was very unstable using the partially frozen VGG19 model over multiple attempts. We could not think of any reasons on why this happened over multiple attempts, but our guess would be the individual clients receive very less training data due to which the aggregate from all the clients is not good enough to generalize on the test data.

E. Conclusion

We observe that ResNet-50 model consistently outperforms VGG-19 model across all experiments. This can be due to two reasons. Firstly, the ResNet-50 model is a much larger model with a greater number of convolution layers as compared to VGG-19. Second, the residual connections in ResNet could help it generalize better.

Moreover, our findings reveal a compelling trend: the comparable performance of these models in both federated and standard training methodologies. This observation underscores the immense promise and potential of federated learning specifically in the realm of medical image analysis. The ability of federated learning to yield similar model performance without compromising data privacy and security signifies its viability as a robust approach to handling sensitive medical data while achieving commendable results.

References

[1] https://pytorch.org/. 3

[2] https://www.kaggle.com/datasets/kondwani/eye-disease-dataset. 3

[3] Rodolfo Stoffel Antunes, Cristiano André da Costa, Arne Küderle, Imrana Abdullahi Yari, and Björn Eskofier. Federated learning for healthcare: Systematic review and architecture proposal. *ACM Trans. Intell. Syst. Technol.*, 13(4), may 2022. 1

[4] Daniel J Beutel, Taner Topal, Akhil Mathur, Xinchi Qiu, Javier Fernandez-Marques, Yan Gao, Lorenzo Sani, Hei Li Kwing, Titouan Parcollet, Pedro PB de Gusmão, and Nicholas D Lane. Flower: A friendly federated learning research framework. *arXiv preprint arXiv:2007.14390*, 2020. 3

[5] Kallista A. Bonawitz, Hubert Eichner, Wolfgang Grieskamp, Dzmitry Huba, Alex Ingerman, Vladimir Ivanov, Chloé Kiddon, Jakub Konečný, Stefano Mazzocchi, H. Brendan McMahan, Timon Van Overveldt, David Petrou, Daniel Ramage, and Jason Roselander. Towards federated learning at scale: System design. *CoRR*, abs/1902.01046, 2019. 1

[6] Ittai Dayan, Holger R Roth, Aoxiao Zhong, Ahmed Harouni, Amilcare Gentili, Anas Z Abidin, Andrew Liu, Anthony Beardsworth Costa, Bradford J Wood, Chien-Sung Tsai, Chih-Hung Wang, Chun-Nan Hsu, C K Lee, Peiying Ruan, Daguang Xu, Dufan Wu, Eddie Huang, Felipe Campos Kitamura, Griffin Lacey, Gustavo César de Antônio Corradi, Gustavo Nino, Hao-Hsin Shin, Hirofumi Obinata, Hui Ren, Jason C Crane, Jesse Tetreault, Jiahui Guan, John W Garrett, Joshua D Kaggie, Jung Gil Park, Keith Dreyer, Krishna Juluru, Kristopher Kersten, Marcio Aloisio Bezerra Cavalcanti Rockenbach, Marius George Linguraru, Masoom A Haider, Meena AbdelMaseeh, Nicola Rieke, Pablo F Damasceno, Pedro Mario Cruz E Silva, Pochuan Wang, Sheng Xu, Shuichi Kawano, Sira Sriswasdi, Soo Young Park, Thomas M Grist, Varun Buch, Watsamon Jantarabenjakul, Weichung Wang, Won Young Tak, Xiang Li, Xihong Lin, Young Joon Kwon, Abood Quraini, Andrew Feng, Andrew N Priest, Baris Turkbey, Benjamin Glicksberg, Bernardo Bizzo, Byung Seok Kim, Carlos Tor-Díez, Chia-Cheng Lee, Chia-Jung Hsu, Chin Lin, Chiu-Ling Lai, Christopher P Hess, Colin Compas, Deepeksha Bhatia, Eric K Oermann, Evan Leibovitz, Hisashi Sasaki, Hitoshi Mori, Isaac Yang, Jae Ho Sohn, Krishna Nand Keshava Murthy, Li-Chen Fu, Matheus Ribeiro Furtado de Mendonça, Mike Fralick, Min Kyu Kang, Mohammad Adil, Natalie Gangai, Peerapon Vateekul, Pierre Elnajjar, Sarah Hickman, Sharmila Majumdar, Shelley L McLeod, Sheridan Reed, Stefan Gräf, Stephanie Harmon, Tatsuya Kodama, Thanyawee Puthanakit, Tony Mazzulli, Vitor Lima de Lavor, Yothin Rakvongthai, Yu Rim Lee, Yuhong Wen, Fiona J Gilbert, Mona G Flores, and Quanzheng Li. Federated learn-

ing for predicting clinical outcomes in patients with COVID-19. *Nat. Med.*, 27(10):1735–1743, Oct. 2021. 2

[7] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *CoRR*, abs/1512.03385, 2015. 2

[8] Heechul Jung, Min-Kook Choi, Jihun Jung, Jin-Hee Lee, Soon Kwon, and Woo Young Jung. Resnet-based vehicle classification and localization in traffic surveillance systems. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, July 2017. 2

[9] Peter Kairouz, H. Brendan McMahan, Brendan Avent, Aurélien Bellet, Mehdi Bennis, Arjun Nitin Bhagoji, Kallista A. Bonawitz, Zachary Charles, Graham Cormode, Rachel Cummings, Rafael G. L. D'Oliveira, Salim El Rouayheb, David Evans, Josh Gardner, Zachary Garrett, Adrià Gascón, Badih Ghazi, Phillip B. Gibbons, Marco Gruteser, Zaïd Harchaoui, Chaoyang He, Lie He, Zhouyuan Huo, Ben Hutchinson, Justin Hsu, Martin Jaggi, Tara Javidi, Gauri Joshi, Mikhail Khodak, Jakub Konečný, Aleksandra Korolova, Farinaz Koushanfar, Sanmi Koyejo, Tancrède Lepoint, Yang Liu, Prateek Mittal, Mehryar Mohri, Richard Nock, Ayfer Özgür, Rasmus Pagh, Mariana Raykova, Hang Qi, Daniel Ramage, Ramesh Raskar, Dawn Song, Weikang Song, Sebastian U. Stich, Ziteng Sun, Ananda Theertha Suresh, Florian Tramèr, Praneeth Vepakomma, Jianyu Wang, Li Xiong, Zheng Xu, Qiang Yang, Felix X. Yu, Han Yu, and Sen Zhao. Advances and open problems in federated learning. *CoRR*, abs/1912.04977, 2019. 1

[10] Haeyun Lee, Young Jun Chai, Hyunjin Joo, Kyungsu Lee, Jae Youn Hwang, Seok-Mo Kim, Kwangsoon Kim, Inn-Chul Nam, June Young Choi, Hyeong Won Yu, Myung-Chul Lee, Hiroo Masuoka, Akira Miyauchi, Kyu Eun Lee, Sungwan Kim, and Hyoun-Joong Kong. Federated learning for thyroid ultrasound image analysis to protect personal information: Validation study in a real health care environment. *JMIR Med. Inform.*, 9(5):e25869, May 2021. 2

[11] H. Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Agüera y Arcas. Communication-efficient learning of deep networks from decentralized data, 2023. 3

[12] H. Brendan McMahan, Eider Moore, Daniel Ramage, and Blaise Agüera y Arcas. Federated learning of deep networks using model averaging. *CoRR*, abs/1602.05629, 2016. 1

[13] Dinh C. Nguyen, Quoc-Viet Pham, Pubudu N. Pathirana, Ming Ding, Aruna Seneviratne, Zihuai Lin, Octavia A. Dobre, and Won-Joo Hwang. Federated learning for smart healthcare: A survey. *CoRR*, abs/2111.08834, 2021. 1

[14] Suraj Rajendran, Jihad S. Obeid, Hamidullah Binol, Ralph D'Agostino, Kristie Foley, Wei Zhang, Philip Austin, Joey Brakefield, Metin N. Gurcan, and Umit Topaloglu. Cloud-based federated learning implementation across medical centers. *JCO Clinical Cancer Informatics*, (5):1–11, 2021. PMID: 33411624. 1

[15] Sashank Reddi, Zachary Charles, Manzil Zaheer, Zachary Garrett, Keith Rush, Jakub Konečný, Sanjiv Kumar, and H. Brendan McMahan. Adaptive federated optimization, 2021. 5

[16] Devvi Sarwinda, Radifa Hilya Paradisa, Alhadi Bustamam, and Pinkie Anggia. Deep learning in image classification using residual network (resnet) variants for detection of colorectal cancer. *Procedia Computer Science*, 179:423–431, 2021. 5th International Conference on Computer Science and Computational Intelligence 2020. 2

[17] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. 2014. 2

[18] Risheng Wang, Tao Lei, Ruixia Cui, Bingtao Zhang, Hongying Meng, and Asoke K. Nandi. Medical image segmentation using deep learning: A survey. *IET Image Processing*, 16(5):1243–1267, Jan. 2022. 1

[19] Wenjian Yao, Jiajun Bai, Wei Liao, Yuheng Chen, Mengjuan Liu, and Yao Xie. From cnn to transformer: A review of medical image segmentation models, 2023. 1