Split Learning for collaborative deep learning in healthcare

Poirot, Maarten G.; Vepakomma, Praneeth; Chang, Ken; Kalpathy-Cramer, Jayashree; Gupta, Rajiv; Raskar, Ramesh

Published: 27/12/2019

Split Learning for collaborative deep learning in healthcare

Maarten G.Poirot[1], **Praneeth Vepakomma**[2], **Ken Chang**[3], **Jayashree Kalpathy-Cramer**[3], **Rajiv Gupta**[1], **Ramesh Raskar**[2]
[1] **Department of Radiology, Massachusetts General Hospital**
maartenpoirot@gmail.com
[2] **Media Lab, Massachusetts Institute of Technology**
vepakom@mit.edu
[3] **Athinoula A. Martinos Center for Biomedical Imaging, Department of Radiology, Massachusetts General Hospital**
kenchang@mit.edu

Abstract

Shortage of labeled data has been holding the surge of deep learning in healthcare back, as sample sizes are often small, patient information cannot be shared openly, and multi-center collaborative studies are a burden to set up. Distributed machine learning methods promise to mitigate these problems. We argue for a split learning based approach and apply this distributed learning method for the first time in the medical field to compare performance against (1) centrally hosted and (2) non collaborative configurations for a range of participants.

Two medical deep learning tasks are used to compare split learning to conventional single and multi center approaches: a binary classification problem of a data set of 9000 fundus photos, and multi-label classification problem of a data set of 156,535 chest X-rays. The several distributed learning setups are compared for a range of 1-50 distributed participants.

Performance of the split learning configuration remained constant for any number of clients compared to a single center study, showing a marked difference compared to the non collaborative configuration after 2 clients ($p < 0.001$) for both sets.

Our results affirm the benefits of collaborative training of deep neural networks in health care. Our work proves the significant benefit of distributed learning in healthcare, and paves the way for future real-world implementations.

1 Introduction

Deep neural networks have become the state-of-the-art for a range of tasks such as image classification, speech recognition, natural language processingCollobert and Weston [2008] and based on complex data such as electronic health records (EHR), imaging, bio-sensors, omics and text.

Learning with these networks relies on vast amounts of structured training data to achieve proper performance such that it increases generalization and robustness Miotto et al. [2017]; Panch et al. [2018]. However, medical sample sizes tend to be small, especially in rarer diseases Dluhos et al. [2017]. Thus traditionally, clinical models have often been trained on small data sets Panch et al. [2018].

Multi-center distributed studies can significantly increase the available amount of data and its diversity by centralization of the data sets, but it comes with several drawbacks: Setting up a multi-center organizational collaboration can be difficult as patient data can often not leave the premise due to ethical or regulatory concerns such as HIPAA Annas [2003]; Mercuri [2004]; Nass et al. [2009];

33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada.

Luxton et al. [2012]. Secondly, institutions might find their data to be too valuable to share Xia et al. [2018]. Lastly, the additional storage and bandwidth required to store this data can be a burden. These factors heavily impede collaboration in health in a traditional setting.

An alternative to centrally hosting information are secure and private distributed learning solutions. These methods include model averagingSu and Chen [2015], large scale synchronous gradient descent (LS-SGD)Chen et al. [2016], federated learningMcMahan et al. [2016], cyclical weight transferChang et al. [2018] and split learning[Gupta and Raskar, 2018; Vepakomma et al., 2017, 2018a; Singh et al., 2019; Sharma et al., 2019]. These models can be compared on several properties, which are performance with respect to a centralized setup, privacy, bandwidth usage and distribution of computational load.

From previous survey by Vepakomma et. al.Vepakomma et al. [2018b] several properties of these methods can be identified and weighed for our clinical implementation. Model averaging and LS-SGD only allow for synchronous training, meaning the model can only continue training after all clients have yielded their input. This would present major logistical challenges, especially when clients work with different network connection speeds or hardware configurations. Other methods like cyclical weight transfer do not preserve optimal performance compared to computing in a centralized setting by design. Lastly, every method differs in the amount of information it reveals. For a more thorough comparison of these methods we would like to refer to the aforementioned papers.

In this study, split learning is applied in the medical field for the first time to our knowledge. Two data sets are used: retinal fundus photos and chest X-rays. For these two data sets performance of a split learning configuration is compared to (1) centrally hosted and (2) non collaborative configuration for a range of number of distributed participants.

2 Related work

The concept of split learning was first introduced by Gupta and Raskar [2018]. In comparisons on the CIFAR 10 and CIFAR 100 data sets; split learning has shown to outperform federated learning and LS-SGD in terms of convergence for accuracy and client side computational requirements Vepakomma et al. [2018a]. In addition, split learning shows improved security by reducing leakage of information as shown by Vepakomma et al Vepakomma et al. [2017].

The paradigm of split learning revolves around splitting up a conventional neural network into several elements that can have different accessibility properties. These elements are '*links*', that together form a '*chain*', making up the full network. The mentioned accessibility properties of these links can either be '*central*', which means they are hosted on the central server location and accessible as black box to all clients, or '*local*', in which case they can only be accessed by the proprietary client.

U-shaped split learning: Although the configuration could potentially take many forms, a particular configuration called the U-shaped configuration Gupta and Raskar [2018]; Vepakomma et al. [2018a] is implemented in this work for its simplicity and suitability for healthcare. This configuration requires no raw data sharing as well as no label sharing. The chain in this configuration consists of three links. As considered from a forward propagation point of view, the first is called '*front*', and is local. It receives raw input data during forward propagation, and returns an obfuscated intermediate representation. The second link is called '*center*' and is centrally hosted. It takes the intermediate representation from the front, and performs most of the computation to return another intermediate representation to the final link called '*back*'. The back is again local and performs the final decoding computation on its input. This local stage is where gradients are computed from the decoded output and labels. This configuration is visualized in figure 1.

When training the model one or more mini-batches can iteratively be forwarded through the chain thereby training both the local, as well as the central links. When training is switched from one client to another, the state of the local links from one clients is downloaded and updated at the next. The system is not dependant on results from all clients to push an update, which resolves the logistical challenges in synchronous training methods mentioned earlier.

Typically but not necessarily, the largest part of trainable networks layers can be found in the central link. This reduces bandwidth used in sharing local states, as well as client side computational cost. This property allows for computation of more complex networks for clients with less computational power, compared to federated learning.

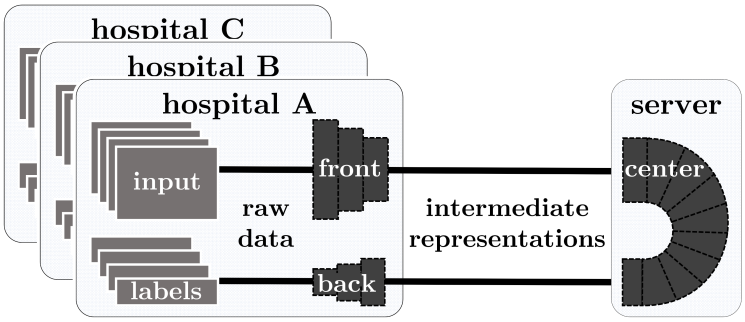

Figure 1: Graphical representation of the U-shaped configuration. Three clients named hospital A, B and C hold their own data and train a collaborative model without sharing raw data.

3 Methods

3.1 Data collection

We utilized the diabetic retinopathy (DR) dataset as previously described in work by Chang et al Chang et al. [2018]. This data set originates from the Kaggle Diabetic Retinopathy dataset Kaggle Inc. [2015] of retinal fundus photos. A subset of 9000 images was used for training and validation to prevent saturation of learning for models when trained non-collaboratively. The original multi-class classification problem was simplified to binary classification of 'normal' and 'abnormal'. Images were downsampled to 256x256 RGB images. For further preprocessing details we refer to Chang et al Chang et al. [2018].

The second data set used was the large chest X-ray dataset '*CheXpert*' Irvin et al. [2019]. The data set consists of 224,316 chest radiographs with labels of 65,240 patients. The problem is posed as a multi-label classification problem of 14 common chest radiographic observations. Cases where labels contained uncertainty were excluded according to the baseline approach as described in the paper. To further decrease the data set size to prevent saturation in non-collaborative setting, some subsets of images of different shapes that were most commonly occurring (320x390 px) were excluded. This resulted in a remaining dataset of 156,535 chest radiographs.

Both data sets were partitioned to cohorts of 75% training, and 25% validation data. When the data sets were further split over multiple clients, training data was split equally. Both partitioning operations are performed randomly without patients overlapping in both cohorts. Validation data was not split so as to retain the validation process even when data is split over many clients.

3.2 Neural networks

For the DR dataset an implementation largely influenced by Chang et al Chang et al. [2018] was employed. A 34-layer residual network (Resnet-34)He et al. [2016] architecture was utilized with Glorot uniform initialization Glorot and Bengio [2010]. Adam Kingma and Ba [2014] optimization using standard parameters ($\beta_1 = 0.9$ and $beta_2 = 0.999$), and default learning rate (10^{-4}) without decay was used. Data was augmented in real-time using random rotations (0-360 degrees) and 50% chance of lateral or axial inversion. Loss was computed used a binary cross entropy loss function. Training was performed on a GeForce GTX TITAN X graphics processing unit until validation accuracy reached a plateau as defined by not decreasing for more than 30 epochs.

For the CheXpert dataset, the implementation as described by Rajpurkar et al Rajpurkar et al. [2017] was used. A 121-layer dense network (DenseNet121) Huang et al. [2016] was pretrained on ImageNetDeng et al. [2009]. Loss was defined computed using a combined sigmoid binary cross entropy loss. Adam optimization using standard parameters ($\beta_1 = 0.9$ and $\beta_2 = 0.999$), and default learning rate (10^{-4}) without decay was used. Batch size used was 24. Data was augmented by 50% chance of lateral inversion. Models were trained until validation loss reached a plateau as defined by not decreasing for more than five epochs. The model with the lowest validation loss was used picked. Training was performed on a Nvidia GeForce GTX 1080 Ti graphics processing unit.

In collaborative mode, every client sequentially trained the network for one epoch. Whenever training switched from one client to the next local client states were copied to next client. In non collaborative mode, a single client was trained on the same sample size of data as it would have had in the collaborative setting.

3.3 Performance analysis

Performance of the DR data set is defined as the highest classification accuracy on the validation set by averaging all clients. For the CheXpert data set, the receiver operating characteristic curves (ROC) were generated from the validation set, for the model state of the client with lowest loss across all mini-batches in the epoch achieving the lowest loss. Final result was the average area under the ROC (AUROC) as shown in Figure 2 across all five competition tasks as defined by the original study (Atelectasis, cardiomegaly, consolidation, edema and pleural effusion).

4 Results

The performance of split learning based configurations is compared to a non collaborative configurations for the DR data set using accuracy, and CheXpert using the AUROC, in figure 2. As shown in the figure, the split learning based approaches on both the CheXpert and diabetic retinopathy datasets performed exceedingly better than performance in non-collaborative settings Irvin et al. [2019]. Experimental results, including bootstrapping results for the diabetic retinopathy set are given in table 1. On the Chexpert split learning dataset mean performance was significantly ($\alpha = 0.005$) lower in non collaborative compared to collaborative setting especially in cases with > 2 clients and two sample two tailed T-test was also used to compare means to reach this conclusion.

5 Discussion and future work

Distributed machine learning based solutions can provide great benefit to the medical field by enhancing seamless collaboration across entities. Split learning has shown benefits compared to alternative distributed learning methods. We have applied split learning in the medical field for the first time and it worked great compared to conventional single and multi-institution setups. Our results show that teaming up in general can give a great performance boost. Our results show that teaming up in a distributed learning setting in general can give a great performance boost in comparison to non-collaboration. In the future we will also compare to federated learning and LS-SGD within the medical setup. These comparisons have already been made recently in the non-medical settings in [Gupta and Raskar, 2018; Vepakomma et al., 2017]. We could investigate alternative weight transfer

Table 1: Performance for diabetic retinopathy

number of clients	Split learning b mean	Non collaborative (C.I.) mean (C.I.)
1	0.888 (0.896, 0.880)	0.869 (0.877, 0.861)
2	0.850 (0.857, 0.843)	0.852 (0.865, 0.839)
3	0.868 (0.875, 0.861)	0.753 (0.766, 0.742)
4	0.884 (0.891, 0.878)	0.754 (0.770, 0.739)
5	0.869 (0.877, 0.861)	0.755 (0.772, 0.738)
8	0.887 (0.894, 0.880)	0.717 (0.733, 0.701)
10	0.858 (0.868, 0.849)	0.676 (0.695, 0.657)
15	0.838 (0.848, 0.829)	0.627 (0.649, 0.603)
20	0.860 (0.868, 0.852)	0.613 (0.632, 0.594)
25	0.850 (0.858, 0.841)	0.607 (0.627, 0.588)
30	0.814 (0.831, 0.797)	0.620 (0.648, 0.590)
35	0.798 (0.819, 0.780)	0.633 (0.656, 0.611)
40	0.852 (0.859, 0.844)	0.595 (0.619, 0.568)
45	0.883 (0.891, 0.876)	0.608 (0.634, 0.581)
50	0.859 (0.869, 0.849)	0.588 (0.611, 0.565)

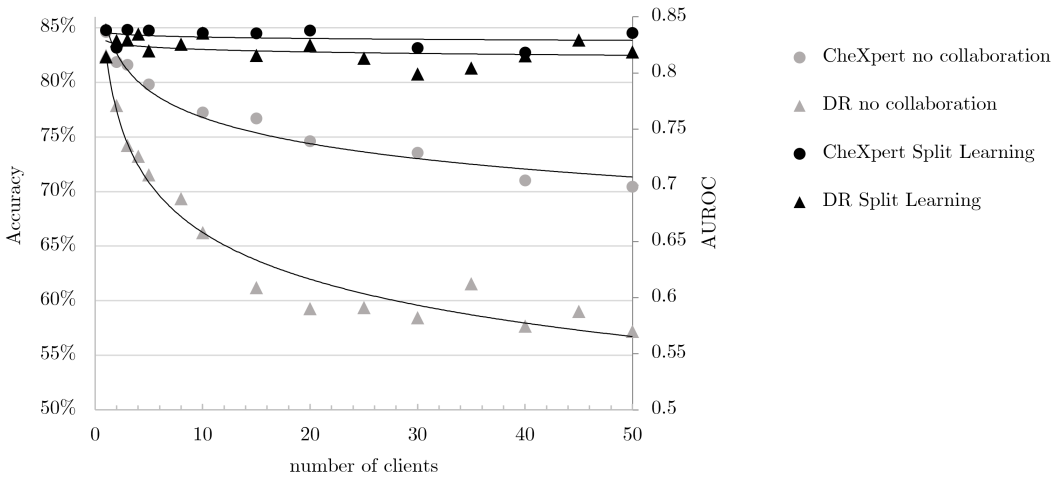

Figure 2: Performance of non collaborative (gray) and split learning (black) configurations. Number of clients refers to the number of clients the data was divided over. As the total amount of data remained constant, this directly relates to each client's sample size.

protocols to aim to improve efficiency. We also plan to investigate privacy enhancements and alternative configurations in Vepakomma et al. [2018a] for healthcare settings via controlled real-world healthcare deployments.

References

George J. Annas. HIPAA regulations - A new era of medical-record privacy? *New England Journal of Medicine*, 348(15):1486–1490, 2003. ISSN 00284793. doi: 10.1056/NEJMlim035027.

Ken Chang, Niranjan Balachandar, Carson Lam, Darvin Yi, James Brown, Andrew Beers, Bruce Rosen, Daniel L Rubin, and Jayashree Kalpathy-Cramer. Distributed deep learning networks among institutions for medical imaging. *Journal of the American Medical Informatics Association : JAMIA*, 25(8):945–954, aug 2018. ISSN 1527-974X (Electronic). doi: 10.1093/jamia/ocy017.

Jianmin Chen, Rajat Monga, Samy Bengio, and Rafal Józefowicz. Revisiting Distributed Synchronous {SGD}. *CoRR*, abs/1604.00981, 2016. URL `http://arxiv.org/abs/1604.00981`.

Ronan Collobert and Jason Weston. A Unified Architecture for Natural Language Processing: Deep Neural Networks with Multitask Learning. In *Proceedings of the 25th International Conference on Machine Learning*, ICML '08, pages 160–167, New York, NY, USA, 2008. ACM. ISBN 978-1-60558-205-4. doi: 10.1145/1390156.1390177. URL `http://doi.acm.org/10.1145/1390156.1390177`.

J Deng, W Dong, R Socher, L Li, Kai Li, and Li Fei-Fei. ImageNet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255, jun 2009. doi: 10.1109/CVPR.2009.5206848.

Petr Dluhos, Daniel Schwarz, Wiepke Cahn, Neeltje van Haren, Rene Kahn, Filip Spaniel, Jiri Horacek, Tomas Kasparek, and Hugo Schnack. Multi-center machine learning in imaging psychiatry: A meta-model approach. *NeuroImage*, 155:10–24, jul 2017. ISSN 1095-9572 (Electronic). doi: 10.1016/j.neuroimage.2017.03.027.

Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feedforward neural networks. In *In Proceedings of the International Conference on Artificial Intelligence and Statistics (AISTATS'10). Society for Artificial Intelligence and Statistics*, 2010.

Otkrist Gupta and Ramesh Raskar. Distributed learning of deep neural network over multiple agents. *Journal of Network and Computer Applications*, 116:1–8, 2018. ISSN 1084-8045. doi: https://doi.org/10.1016/j.jnca.2018.05.003. URL `http://www.sciencedirect.com/science/article/pii/S1084804518301590`.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016.

Gao Huang, Zhuang Liu, and Kilian Q Weinberger. Densely Connected Convolutional Networks. *CoRR*, abs/1608.06993, 2016. URL http://arxiv.org/abs/1608.06993.

Jeremy Irvin, Pranav Rajpurkar, Michael Ko, Yifan Yu, Silviana Ciurea-Ilcus, Chris Chute, Henrik Marklund, Behzad Haghgoo, Robyn L. Ball, Katie S. Shpanskaya, Jayne Seekins, David A. Mong, Safwan S. Halabi, Jesse K. Sandberg, Ricky Jones, David B. Larson, Curtis P. Langlotz, Bhavik N. Patel, Matthew P. Lungren, and Andrew Y. Ng. Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison. *CoRR*, abs/1901.07031, 2019. URL http://arxiv.org/abs/1901.07031.

Kaggle Inc. Diabetic Retinopathy Detection — Kaggle, 2015. URL https://www.kaggle.com/c/diabetic-retinopathy-detection.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

David D Luxton, Robert A Kayl, and Matthew C Mishkind. mHealth data security: the need for HIPAA-compliant standardization. *Telemedicine journal and e-health : the official journal of the American Telemedicine Association*, 18(4):284–288, may 2012. ISSN 1556-3669 (Electronic). doi: 10.1089/tmj.2011.0180.

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. URL http://arxiv.org/abs/1602.05629.

Rebecca T Mercuri. The HIPAA-potamus in Health Care Data Security. *Commun. ACM*, 47(7): 25–28, jul 2004. ISSN 0001-0782. doi: 10.1145/1005817.1005840. URL http://doi.acm.org/10.1145/1005817.1005840.

Riccardo Miotto, Fei Wang, Shuang Wang, Xiaoqian Jiang, and Joel T Dudley. Deep learning for healthcare: review, opportunities and challenges. *Briefings in Bioinformatics*, 19(6):1236–1246, 2017. ISSN 1477-4054. doi: 10.1093/bib/bbx044. URL https://doi.org/10.1093/bib/bbx044.

Sharyl J. Nass, Laura A. Levit, and Lawrence O. Gostin. *Beyond the HIPAA Privacy Rule: Enhancing Privacy, Improving Health Through Research*. National Academies Press, 2009. ISBN 9780309141376. doi: 10.17226/12458.

Trishan Panch, Peter Szolovits, and Rifat Atun. Artificial intelligence, machine learning and health systems. *Journal of global health*, 8(2):20303, dec 2018. ISSN 2047-2986 (Electronic). doi: 10.7189/jogh.08.020303.

Pranav Rajpurkar, Jeremy Irvin, Kaylie Zhu, Brandon Yang, Hershel Mehta, Tony Duan, Daisy Yi Ding, Aarti Bagul, Curtis Langlotz, Katie S. Shpanskaya, Matthew P. Lungren, and Andrew Y. Ng. Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. *CoRR*, abs/1711.05225, 2017. URL http://arxiv.org/abs/1711.05225.

Vivek Sharma, Praneeth Vepakomma, Tristan Swedish, Ken Chang, Jayashree Kalpathy-Cramer, and Ramesh Raskar. Expertmatcher: Automating ml model selection for clients using hidden representations. *arXiv preprint arXiv:1910.03731*, 2019.

Abhishek Singh, Praneeth Vepakomma, Otkrist Gupta, and Ramesh Raskar. Detailed comparison of communication efficiency of split learning and federated learning. *arXiv preprint arXiv:1909.09145*, 2019.

Hang Su and Haoyu Chen. Experiments on parallel training of deep neural network using model averaging. *CoRR*, abs/1507.01239, 2015. URL http://arxiv.org/abs/1507.01239.

Praneeth Vepakomma, Otkrist Gupta, Abhimanyu Dubey, and Ramesh Raskar. Reducing leakage in distributed deep learning. *AI for social good*, pages 1–6, 2017.

Praneeth Vepakomma, Otkrist Gupta, Tristan Swedish, and Ramesh Raskar. Split learning for health: Distributed deep learning without sharing raw patient data. *CoRR*, abs/1812.00564, 2018a. URL http://arxiv.org/abs/1812.00564.

Praneeth Vepakomma, Tristan Swedish, Ramesh Raskar, Otkrist Gupta, and Abhimanyu Dubey. No peek: A survey of private distributed deep learning. *CoRR*, abs/1812.03288, 2018b. URL http://arxiv.org/abs/1812.03288.

Weiyi Xia, Zhiyu Wan, Zhijun Yin, James Gaupp, Yongtai Liu, Ellen Wright Clayton, Murat Kantarcioglu, Yevgeniy Vorobeychik, and Bradley A Malin. It's all in the timing: calibrating temporal penalties for biomedical data sharing. *Journal of the American Medical Informatics Association : JAMIA*, 25(1):25–31, jan 2018. ISSN 1527-974X (Electronic). doi: 10.1093/jamia/ocx101.

References

George J. Annas. HIPAA regulations - A new era of medical-record privacy? *New England Journal of Medicine*, 348(15):1486–1490, 2003. ISSN 00284793. doi: 10.1056/NEJMlim035027.

Ken Chang, Niranjan Balachandar, Carson Lam, Darvin Yi, James Brown, Andrew Beers, Bruce Rosen, Daniel L Rubin, and Jayashree Kalpathy-Cramer. Distributed deep learning networks among institutions for medical imaging. *Journal of the American Medical Informatics Association : JAMIA*, 25(8):945–954, aug 2018. ISSN 1527-974X (Electronic). doi: 10.1093/jamia/ocy017.

Jianmin Chen, Rajat Monga, Samy Bengio, and Rafal Józefowicz. Revisiting Distributed Synchronous {SGD}. *CoRR*, abs/1604.00981, 2016. URL http://arxiv.org/abs/1604.00981.

Ronan Collobert and Jason Weston. A Unified Architecture for Natural Language Processing: Deep Neural Networks with Multitask Learning. In *Proceedings of the 25th International Conference on Machine Learning*, ICML '08, pages 160–167, New York, NY, USA, 2008. ACM. ISBN 978-1-60558-205-4. doi: 10.1145/1390156.1390177. URL http://doi.acm.org/10.1145/1390156.1390177.

J Deng, W Dong, R Socher, L Li, Kai Li, and Li Fei-Fei. ImageNet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255, jun 2009. doi: 10.1109/CVPR.2009.5206848.

Petr Dluhos, Daniel Schwarz, Wiepke Cahn, Neeltje van Haren, Rene Kahn, Filip Spaniel, Jiri Horacek, Tomas Kasparek, and Hugo Schnack. Multi-center machine learning in imaging psychiatry: A meta-model approach. *NeuroImage*, 155:10–24, jul 2017. ISSN 1095-9572 (Electronic). doi: 10.1016/j.neuroimage.2017.03.027.

Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feedforward neural networks. In *In Proceedings of the International Conference on Artificial Intelligence and Statistics (AISTATS'10). Society for Artificial Intelligence and Statistics*, 2010.

Otkrist Gupta and Ramesh Raskar. Distributed learning of deep neural network over multiple agents. *Journal of Network and Computer Applications*, 116:1–8, 2018. ISSN 1084-8045. doi: https://doi.org/10.1016/j.jnca.2018.05.003. URL http://www.sciencedirect.com/science/article/pii/S1084804518301590.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016.

Gao Huang, Zhuang Liu, and Kilian Q Weinberger. Densely Connected Convolutional Networks. *CoRR*, abs/1608.06993, 2016. URL http://arxiv.org/abs/1608.06993.

Jeremy Irvin, Pranav Rajpurkar, Michael Ko, Yifan Yu, Silviana Ciurea-Ilcus, Chris Chute, Henrik Marklund, Behzad Haghgoo, Robyn L. Ball, Katie S. Shpanskaya, Jayne Seekins, David A. Mong, Safwan S. Halabi, Jesse K. Sandberg, Ricky Jones, David B. Larson, Curtis P. Langlotz, Bhavik N. Patel, Matthew P. Lungren, and Andrew Y. Ng. Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison. *CoRR*, abs/1901.07031, 2019. URL http://arxiv.org/abs/1901.07031.

Kaggle Inc. Diabetic Retinopathy Detection — Kaggle, 2015. URL https://www.kaggle.com/c/diabetic-retinopathy-detection.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

David D Luxton, Robert A Kayl, and Matthew C Mishkind. mHealth data security: the need for HIPAA-compliant standardization. *Telemedicine journal and e-health : the official journal of the American Telemedicine Association*, 18(4):284–288, may 2012. ISSN 1556-3669 (Electronic). doi: 10.1089/tmj.2011.0180.

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. URL http://arxiv.org/abs/1602.05629.

Rebecca T Mercuri. The HIPAA-potamus in Health Care Data Security. *Commun. ACM*, 47(7): 25–28, jul 2004. ISSN 0001-0782. doi: 10.1145/1005817.1005840. URL http://doi.acm.org/10.1145/1005817.1005840.

Riccardo Miotto, Fei Wang, Shuang Wang, Xiaoqian Jiang, and Joel T Dudley. Deep learning for healthcare: review, opportunities and challenges. *Briefings in Bioinformatics*, 19(6):1236–1246, 2017. ISSN 1477-4054. doi: 10.1093/bib/bbx044. URL https://doi.org/10.1093/bib/bbx044.

Sharyl J. Nass, Laura A. Levit, and Lawrence O. Gostin. *Beyond the HIPAA Privacy Rule: Enhancing Privacy, Improving Health Through Research*. National Academies Press, 2009. ISBN 9780309141376. doi: 10.17226/12458.

Trishan Panch, Peter Szolovits, and Rifat Atun. Artificial intelligence, machine learning and health systems. *Journal of global health*, 8(2):20303, dec 2018. ISSN 2047-2986 (Electronic). doi: 10.7189/jogh.08.020303.

Pranav Rajpurkar, Jeremy Irvin, Kaylie Zhu, Brandon Yang, Hershel Mehta, Tony Duan, Daisy Yi Ding, Aarti Bagul, Curtis Langlotz, Katie S. Shpanskaya, Matthew P. Lungren, and Andrew Y. Ng. Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. *CoRR*, abs/1711.05225, 2017. URL http://arxiv.org/abs/1711.05225.

Vivek Sharma, Praneeth Vepakomma, Tristan Swedish, Ken Chang, Jayashree Kalpathy-Cramer, and Ramesh Raskar. Expertmatcher: Automating ml model selection for clients using hidden representations. *arXiv preprint arXiv:1910.03731*, 2019.

Abhishek Singh, Praneeth Vepakomma, Otkrist Gupta, and Ramesh Raskar. Detailed comparison of communication efficiency of split learning and federated learning. *arXiv preprint arXiv:1909.09145*, 2019.

Hang Su and Haoyu Chen. Experiments on parallel training of deep neural network using model averaging. *CoRR*, abs/1507.01239, 2015. URL http://arxiv.org/abs/1507.01239.

Praneeth Vepakomma, Otkrist Gupta, Abhimanyu Dubey, and Ramesh Raskar. Reducing leakage in distributed deep learning. *AI for social good*, pages 1–6, 2017.

Praneeth Vepakomma, Otkrist Gupta, Tristan Swedish, and Ramesh Raskar. Split learning for health: Distributed deep learning without sharing raw patient data. *CoRR*, abs/1812.00564, 2018a. URL http://arxiv.org/abs/1812.00564.

Praneeth Vepakomma, Tristan Swedish, Ramesh Raskar, Otkrist Gupta, and Abhimanyu Dubey. No peek: A survey of private distributed deep learning. *CoRR*, abs/1812.03288, 2018b. URL http://arxiv.org/abs/1812.03288.

Weiyi Xia, Zhiyu Wan, Zhijun Yin, James Gaupp, Yongtai Liu, Ellen Wright Clayton, Murat Kantarcioglu, Yevgeniy Vorobeychik, and Bradley A Malin. It's all in the timing: calibrating temporal penalties for biomedical data sharing. *Journal of the American Medical Informatics Association : JAMIA*, 25(1):25–31, jan 2018. ISSN 1527-974X (Electronic). doi: 10.1093/jamia/ocx101.