Federated learning for heart segmentation

Thibaud Misonne
Louvain School of Engineering,
Louvain-la-Neuve, Belgium
Email: thibaud.misonne@student.uclouvain.be

Sébastien Jodogne
Institute of Information and Communication Technologies,
Electronics and Applied Mathematics,
Louvain-la-Neuve, Belgium
Email: sebastien.jodogne@uclouvain.be

***Abstract*—Training deep learning models for medical imaging requires access to large volumes of sensitive patient data. To this end, the models are generally trained on centralized, de-identified databases that are hard to collect because of privacy requirements. Federated learning proposes an alternative approach, in which a coalition of hospitals collaboratively trains a central model without exchanging any clinical data. This paper explores the combination of federated learning with U-Net models, and applies it to the task of image segmentation of the heart. A variant of federated learning referred to as *federated equal-chances* that improves segmentation performance on unbalanced datasets is introduced as well.**

I. Introduction

Artificial intelligence is being increasingly applied to medical imaging. In particular, the task of the automated segmentation of organs or biological structures has attracted a lot of interest. The family of the U-Net deep learning models has proved to be highly successful on bioimaging segmentation tasks [1]. Radiotherapy offers a great opportunity to study such segmentation algorithms in the context of human radiology. Indeed, the oncologists working in any radiotherapy department all around the world have to delineate organs-at-risk and tumors so as to propose efficient, highly personalized treatments to their patients [2]. This results in radiotherapy departments gathering large volumes of labeled images that could be extremely useful to train artificial intelligence algorithms.

However, such labeled medical images are not easily accessible to researchers, because they correspond to clinical data that must be carefully protected, and that should not be communicated outside of the hospitals without patient's consent. As a consequence, even though free and open-source software are widely available to extract DICOM images from the Picture Archiving and Communication System (PACS) of the hospital, to de-identify them, then to send them to external artificial intelligence experts [3, 4], such clinical trials necessitate the setup of complex research protocols because they typically associate two distinct entities (one hospital and one university). Some large datasets have nevertheless been released as open data, notably thanks to the efforts of *The Cancer Imaging Archive* project [5], but this is a small fraction of all the data that is produced by hospitals on a daily basis.

An alternative approach would be to train deep learning models directly inside the hospitals, without having to communicate patient data to the outside world. In practice however, training a complex segmentation model requires a lot of annotated data, and the amount of data available inside one single hospital may not be sufficient to effectively train the model. Furthermore, few hospitals employ data scientists who are granted to access patient data in order to do in-house artificial intelligence research, which limits the applicability of this approach to a handful of university hospitals.

Federated learning might be a game changer to train deep learning models for healthcare [6]. In the typical federated learning setup, many mobile devices collaborate to train a shared model in a decentralized way, without ever communicating personal data to protect privacy [7]. Federated learning has originally been motivated by supervised learning applications characterized by four key properties: non-i.i.d. (non independent and identically distributed) data [8], unbalanced data, massively distributed data, and limited communications.

Federated learning allows to envision the training of a central deep learning model for medical applications by a coalition of several hospitals and research teams, while avoiding any communication of patient data between these entities. Figure 1 transposes the typical setup of federated learning to the context of medical imaging [9–11]. The imaging data tends to be non-i.i.d. between hospitals because of different clinical teams and possibly different medical guidelines. Note that even within one single hospital, data related to image segmentation tasks might possibly be non-i.i.d. because of intra-observer and inter-observer variability [12]. Furthermore, datasets can be strongly unbalanced between the hospitals of the coalition, typically if smaller hospitals collaborate with larger hospitals. Summarizing, similarly to the original federated learning setup, the non-i.i.d. and unbalanced key properties of the data hold. However, the size of a healthcare coalition is much smaller than a coalition involving mobile devices. In addition, most hospitals benefit from good network connectivity. This implies that the two last key properties of original federated learning (i.e. massively distributed data and limited communications) are not necessarily fulfilled, which hopefully implies that learning is facilitated.

According to this discussion, this paper studies the combination of U-Net architectures with federated learning, and applies it to the segmentation of the heart on CT-scan images. Experimental results compare the predictions of the classical, centralized learning strategy with a federated learning strategy.

II. Federated learning for U-Nets

In this section, it is explained how U-Nets can be used in conjunction with federated learning.

A. U-Net architecture

U-Net is a highly successful deep learning architecture for bioimaging segmentation [1]. Contrarily to the sliding-window approach, in which all the possible patches of a

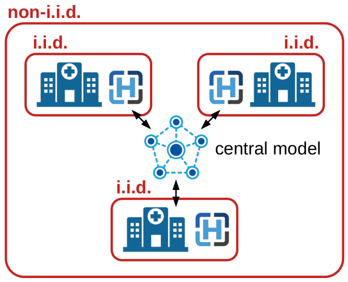


Fig. 1: Federated learning in the context of medical imaging. Several hospitals contribute to the training of one single deep learning model. No patient data leaves any hospital at any time: Only the parameters of the central model are exchanged.

fixed size in an image are scanned by a convolutional neural network producing one binary output for each patch [13], U-Nets consider the input image as a whole. U-Nets first apply a contraction path to generate high-level features, then apply an expansion path to output the segmentation mask, resulting in the typical "U" shape that is depicted in Figure 2. This architecture allows the global spatial structure of the medical images to be taken into account. Generating a segmentation mask given an input image is also much faster than using the sliding-window approach.

In this work, the U-Net models are trained on the individual axial slices of segmented CT-scan images so as to minimize a smoothed version of the 2D Dice loss function [14], that is defined as:

$$\mathcal{L}(\hat{Y}, Y) = -\frac{2\left(\sum_{(x,y)} \hat{Y}(x,y)Y(x,y)\right) + \epsilon}{\left(\sum_{(x,y)} \hat{Y}(x,y)\right) + \left(\sum_{(x,y)} Y(x,y)\right) + \epsilon},$$

where $Y(x,y) \in \{0,1\}$ (resp. $\hat{Y}(x,y)$) denotes the expected binary value (resp. the actual prediction of the U-Net) of the pixel at coordinates (x,y), and where ϵ is a coefficient that prevents divisions by zeros as well as vanishing gradients induced by the slices that do not contain any pixel belonging to their segmentation mask.

B. Federated averaging

Federated averaging is a well-known, effective algorithm for federated learning [7]. In this algorithm, the weights of the central deep learning model are collaboratively optimized by a swarm of clients, each of them using their own private dataset.

Federated averaging is divided in a number of rounds. At each round t, federated averaging randomly selects a subset of K clients, sends them the current parameters w_t of the central model, asks the selected K clients to optimize the parameters on their own local dataset during a fixed number of epochs, then collects back the parameters $w_{t+1}^{(k)}$ as independently optimized by each client k. The parameters of the central model are then updated by taking the weighted average of the locally optimized parameters as follows:

$$w_{t+1} \leftarrow \sum_{k=1}^{K} \frac{n_k}{n} w_{t+1}^{(k)}, \tag{1}$$

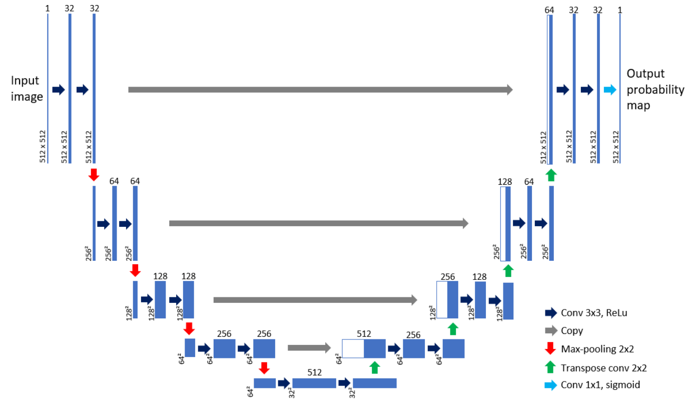


Fig. 2: The U-Net architecture that is used throughout this work. The main differences with respect to the original U-Net architecture [1] are the size of the input and output images (both 512×512 pixels), the number of filters used for the convolutions (it ranges from 32 to 512 filters), and the zero-padding that is used in convolutions (which preserves the image sizes throughout the network). The resulting U-Net models have exactly 7,759,521 floating-point parameters.

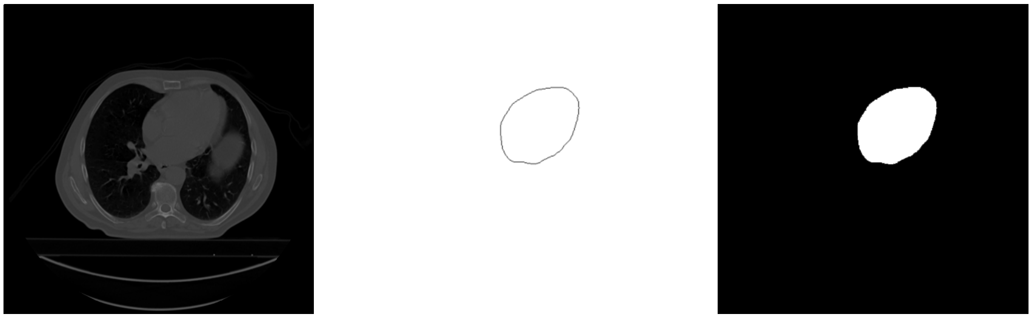
where n_k is the size of the local dataset of client k, and where $n = \sum_{k=1}^{K} n_k$ is the sum of the sizes of all the local datasets of the clients selected at round t.

In this work, the clients correspond to the hospitals. Because a typical learning coalition for healthcare will only contain a handful of hospitals, there is no random selection of the clients: All the hospitals optimize the model at each round t. Furthermore, this paper applies federated averaging to U-Nets: The weights w_t thus correspond to the 7 millions of floating-point parameters of our U-Nets. Because each parameter is encoded as a 32-bits floating-point number in our setup, exchanging one full U-Net model through a computer network requires around 30MB of data. This is not an issue in practice because hospitals largely feature good network connectivity, and because the time that is needed for one round of U-Net optimization is typically much higher than for the transmission of data.

C. Equal chances through data augmentation

The classical federated averaging algorithm described in Section II-B can be problematic if the local datasets of the different hospitals are significantly unbalanced. Indeed, Equation 1 gives a much stronger importance to the model updates that stem from the larger datasets, hereby affecting the quality of the final model when tested on the smaller datasets.

As a consequence, we propose to leverage *data augmentation* so that every client in the coalition uses a training set of equal size. Data augmentation is commonly used to prevent overfitting of deep learning models in computer vision [15]: It consists in artificially increasing the size of a dataset using label-preserving transformations, such as rotations, flips or zooms. Let m_k denote the size of the original local dataset of the client k, and let M be the size of the largest dataset. The *federated equal-chances* algorithm consists, for each client k at each round t, in replacing the m_k original items in the



(a) (b) (c)

Fig. 3: Example of one typical entry in the datasets: (a) the 2D slice from the CT-scan image, (b) the contour of the heart in this slice, as delineated by the physician and encoded as a sequence of points defining a closed polygon, and (c) the binary segmentation mask that is obtained by filling up the polygon using a standard scan-line algorithm.

dataset, by $n_k = M$ new items obtained through online data augmentation. In this flavor of federated averaging, $n = KM$ and Equation 1 becomes an unweighted average of the updates.

III. Experimental results

The federated averaging and federated equal-chances algorithms are now applied to the task of heart segmentation on CT-scan images using U-Nets.

A. Datasets

Our experiments emulate an environment similar to the one depicted in Figure 1. The three hospitals in the figure are replaced by three unrelated datasets provided by *The Cancer Imaging Archive* project [5]. The three considered datasets are NSCLC-Radiomics [16, 17], Pediatric-CT-SEG [18, 19] and Lung CT Segmentation Challenge 2017 (LCTSC) [20, 21]. They provide DICOM CT-scan images together with the delineation of the heart encoded as DICOM RT-STRUCT instances, as routinely used in radiotherapy and nuclear medicine departments. Figure 3 illustrates a typical entry in these datasets.

Because these three datasets have been gathered and delineated separately by different research teams for different purposes, they provide non-i.i.d. distributions by design. Furthermore, they have varying sizes: NSCLC-Radiomics, Pediatric-CT-SEG and LCTSC respectively contain 127, 349 and 59 patients with a valid segmentation of the heart. These datasets are thus representative of a federated learning setup with non-i.i.d. and unbalanced data. Also, the three considered datasets contain images of whole chests, many axial slices of which do not include the heart. The datasets have thus been pre-processed in order to remove the axial slices that are farther than 50% the height of the heart, with respect to the top and bottom slices of the heart. The pre-processed datasets therefore contain the same number of slices including the heart than slices not including it.

Each dataset is randomly split into three independent parts: a train, a validation and a test set, containing respectively 70%, 20% and 10% of the patients. The validation set is used for the early stopping of the training to avoid overfitting. Importantly, the train/validation/test sets are defined at the patient level, not at the slice level. Working at the slice level would totally bias the results, as this would induce a strong correlation between the train/validation/test sets.

B. Learning parameters and evaluation metrics

One single 3D segmentation of the heart defines multiple binary segmentation masks, one for each 2D axial slice of the CT-scan image: The U-Net is trained to map those individual 2D axial slices to their associated binary segmentation masks. However, the performance metrics related to organ segmentation should consider the heart as a whole in the 3D space. This motivates the introduction of a 3D version of Dice similarity coefficient defined as follows [22]:

$$\text{DSC}_{\text{3D}}(\hat{Y}, Y) = \frac{2\left(\sum\limits_{(x,y,z)} \hat{Y}(x,y,z) Y(x,y,z)\right)}{\left(\sum\limits_{(x,y,z)} \hat{Y}(x,y,z)\right) + \left(\sum\limits_{(x,y,z)} Y(x,y,z)\right)},$$

where $Y(x,y,z) \in \{0,1\}$ (resp. $\hat{Y}(x,y,z)$) denotes the expected binary value (resp. the prediction of the U-Net) of the voxel (x,y) in the axial slice whose coordinate in the 3D stack is z. This 3D performance metrics was used to assess the final quality of the models.

The source code was written in Python using Keras with the TensorFlow back-end. The Adam optimizer was used, with random initial weights, a learning rate of $5 \cdot 10^{-5}$ and a batch size of 1. Note that this batch size of 1 corresponds to the value that was used in the original U-Net paper [1]. Dropout layers were introduced to prevent overfitting [23]. Online data augmentation was applied through Keras data generators, which means that the augmented images were different at each epoch. The following transformations were experimentally found to provide best results for data augmentation: rotations between $-25°$ and $25°$, zoom-in and zoom-out up to 8%, and slight modifications of the Hounsfield values of the CT-scan between -1.5% and 1.5%. It was found that using these parameters led to an improvement of about 1% in the Dice loss function if training the central model.

The learning was divided in two phases: Firstly, the U-Nets were trained only on the slices that contain a part of the heart up to convergence, with $\epsilon = 1$ in the loss function.

TABLE I: Value of DSC_{3D} for the different strategies

	NSCLC *127 patients*	Pediatric-CT-SEG *349 patients*	LCTSC *59 patients*
Centralized learning	*0.884*	*0.945*	*0.900*
Local dataset only	0.858	0.939	0.901
Federated averaging	0.867	**0.943**	0.891
Federated equal-chances	**0.879**	0.940	**0.922**

TABLE II: Approximate learning time

Centralized learning (535 patients)	9 *hours*
Local NSCLC-Radiomics *(127 patients)*	0.8 hour
Local Pediatric-CT-SEG *(349 patients)*	5.3 hours
Local LCTSC *(59 patients)*	0.5 hour
Federated averaging (accumulated time)	18.3 hours
Federated equal-chances (accumulated time)	30.5 hours

Secondly, the U-Nets were refined by adding the slices that do not contain the heart, with $\epsilon = 5000$. Without this two-phases approach, the learning process tends to diverge.

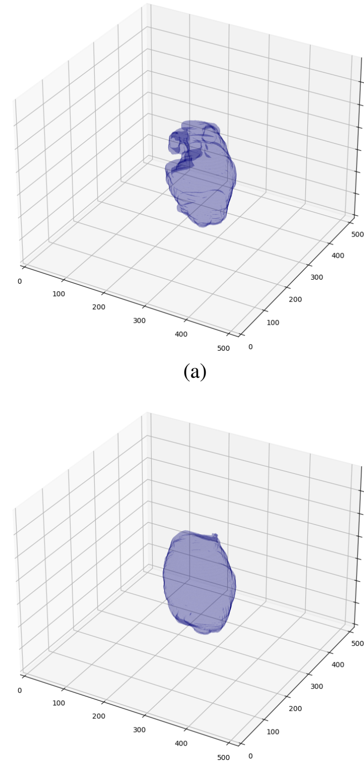
C. Results and discussion

U-Net models were trained using the four strategies presented in this paper. The obtained values for the 3D Dice similarity coefficient are reported in Table I. In this table, the centralized learning is the traditional, reference scenario in which the training is done on one central database that gathers the three datasets. Table II reports the time that is necessary to learn these different models on an Intel Xeon 6244 server equipped with a NVIDIA RTX 2080 Ti GPU.

The first interesting result is that the federated equal-chances algorithm systematically provides better scores than if the model was trained only on one of the three considered datasets. In the real world, this confirms the intuition that a hospital with a limited dataset should consider joining a coalition of hospitals. Table I also indicates that the federated equal-chances algorithm obtains better results than the standard federated averaging algorithm on the smaller datasets, while not being significantly detrimental to the larger dataset.

As expected, centralized learning provides slightly better scores in general, except on the LCTSC dataset, while having an accumulated runtime that is lower than the federated learning algorithms. However, as explained in the Introduction, centralized learning comes with the difficulty of gathering de-identified data outside of hospitals, which may or may not be a viable option depending on the real-world scenario. Furthermore, the reported runtimes for federated learning are accumulated, as though the trainings were done sequentially by the three hospitals. In practice, these trainings would be done concurrently, which would divide the total runtime.

Finally, it is interesting to note that NSCLC-Radiomics has a lower score than the two other datasets for all the strategies. The explanation is provided in Figure 4: It appears that there is an inter-observer variability were the contours of the heart have not be delineated using the same conventions throughout the dataset, which results in lower score on the LCTSC dataset in Table I. This however does not prevent the U-Nets from converging and providing valuable segmentations.



(a) (b)

Fig. 4: 3D rendering of the segmentation of the heart in one DICOM study taken from the NSCLC-Radiomics dataset that illustrates the non-i.i.d. problem: (a) contour delineated by the physician, (b) contour segmented by the U-Net model. The difference in the results is explained by the fact that this particular physician has not followed the same conventions for delineating the heart than other physicians, including many more details in the contours. This is an example of inter-observer variability [12].

IV. Conclusion

In this paper, the combination of federated learning with U-Net models was applied to the task of heart segmentation in CT-scan images. A new variant of the federated averaging algorithm referred to as *federated equal-chances* was introduced to improve the segmentation performance by balancing the size of the datasets through data augmentation. Experimental results indicate that federated learning by a coalition of hospitals is a viable alternative to the traditional centralized learning on a global database gathering patient data from different hospitals.

Future work will apply federated learning to other bioimaging tasks, notably the segmentation of the lungs and of the esophagus, and to other deep learning models than U-Nets [24]. From a software engineering point of view, it is planned to develop a framework to actually deploy federated learning inside clinical environments as a plugin to the Orthanc server [4]. This development will require to setup a network infrastructure in which the hospitals connect as clients to a central server that manages the global model, which necessitates a focus on the security of the communications, on data management, and on the orchestration of data augmentation [25].

Acknowledgments

Computational resources have been provided by the supercomputing facilities of the *Université catholique de Louvain* (CISM/UCL) and the *Consortium des Équipements de Calcul Intensif en Fédération Wallonie Bruxelles* (CÉCI) funded by the *Fonds de la Recherche Scientifique de Belgique* (F.R.S.-FNRS) under convention 2.5020.11 and by the Walloon Region. The authors also wish to thank Eliott Brion and Jean Léger for their helpful insights regarding U-Net architectures.

References

[1] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Springer International Publishing, 2015, pp. 234–241.

[2] C. Sawyers, "Targeted cancer therapy," *Nature*, vol. 432, no. 7015, pp. 294–297, Nov. 2004. [Online]. Available: https://doi.org/10.1038/nature03095

[3] National Electrical Manufacturers Association, "ISO 12052 / NEMA PS3, Digital Imaging and Communications in Medicine (DICOM) standard," Rosslyn, VA, USA, 2022. [Online]. Available: http://www.dicomstandard.org/

[4] S. Jodogne, "The Orthanc ecosystem for medical imaging," *Journal of Digital Imaging*, vol. 31, no. 3, pp. 341–352, 2018. [Online]. Available: http://hdl.handle.net/2078.1/257255

[5] K. Clark, B. Vendt, K. Smith, J. Freymann, J. Kirby, P. Koppel, S. Moore, S. Phillips, D. Maffitt, M. Pringle, L. Tarbox, and F. Prior, "The Cancer Imaging Archive (TCIA): Maintaining and operating a public information repository," *Journal of Digital Imaging*, vol. 26, no. 6, pp. 1045–1057, Jul. 2013. [Online]. Available: https://doi.org/10.1007/s10278-013-9622-7

[6] N. Rieke, J. Hancox, W. Li, F. Milletarì, H. R. Roth, S. Albarqouni, S. Bakas, M. N. Galtier, B. A. Landman, K. Maier-Hein, S. Ourselin, M. Sheller, R. M. Summers, A. Trask, D. Xu, M. Baust, and M. J. Cardoso, "The future of digital health with federated learning," *npj Digital Medicine*, vol. 3, no. 1, Sep. 2020. [Online]. Available: https://doi.org/10.1038/s41746-020-00323-1

[7] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y. Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, ser. Proceedings of Machine Learning Research, A. Singh and J. Zhu, Eds., vol. 54. PMLR, 20–22 Apr 2017, pp. 1273–1282. [Online]. Available: https://proceedings.mlr.press/v54/mcmahan17a.html

[8] H. Zhu, J. Xu, S. Liu, and Y. Jin, "Federated learning on non-IID data: A survey," *Neurocomputing*, vol. 465, pp. 371–390, Nov. 2021. [Online]. Available: https://doi.org/10.1016/j.neucom.2021.07.098

[9] M. J. Sheller, G. A. Reina, B. Edwards, J. Martin, and S. Bakas, "Multi-institutional deep learning modeling without sharing patient data: A feasibility study on brain tumor segmentation," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. Springer International Publishing, 2019, pp. 92–104. [Online]. Available: https://doi.org/10.1007/978-3-030-11723-8_9

[10] A. G. Roy, S. Siddiqui, S. Pölsterl, N. Navab, and C. Wachinger, "BrainTorrent: A peer-to-peer environment for decentralized federated learning," *arXiv preprint arXiv:1905.06731*, 2019.

[11] X. Li, Y. Gu, N. Dvornek, L. H. Staib, P. Ventola, and J. S. Duncan, "Multi-site fMRI analysis using privacy-preserving federated learning and domain adaptation: ABIDE results," *Medical Image Analysis*, vol. 65, p. 101765, Oct. 2020. [Online]. Available: https://doi.org/10.1016/j.media.2020.101765

[12] C. Fiorino, M. Reni, A. Bolognesi, G. M. Cattaneo, and R. Calandrino, "Intra- and inter-observer variability in contouring prostate and seminal vesicles: Implications for conformal treatment planning," *Radiotherapy and Oncology*, vol. 47, no. 3, pp. 285–292, Jun. 1998. [Online]. Available: https://doi.org/10.1016/s0167-8140(98)00021-8

[13] D. Ciresan, A. Giusti, L. Gambardella, and J. Schmidhuber, "Deep neural networks segment neuronal membranes in electron microscopy images," in *Advances in Neural Information Processing Systems*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds., vol. 25. Curran Associates, Inc., 2012. [Online]. Available: http://papers.nips.cc/paper/by-source-2012-1292

[14] C. H. Sudre, W. Li, T. Vercauteren, S. Ourselin, and M. J. Cardoso, "Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Springer International Publishing, 2017, pp. 240–248. [Online]. Available: https://doi.org/10.1007/978-3-319-67558-9_28

[15] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of Big Data*, vol. 6, no. 1, Jul. 2019. [Online]. Available: https://doi.org/10.1186/s40537-019-0197-0

[16] H. J. W. L. Aerts, E. R. Velazquez, R. T. H. Leijenaar, C. Parmar, P. Grossmann, S. Carvalho, J. Bussink, R. Monshouwer, B. Haibe-Kains, D. Rietveld, F. Hoebers, M. M. Rietbergen, C. R. Leemans, A. Dekker, J. Quackenbush, R. J. Gillies, and P. Lambin, "Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach," *Nature Communications*, vol. 5, no. 1, Jun. 2014. [Online]. Available: https://doi.org/10.1038/ncomms5006

[17] H. J. W. L. Aerts, L. Wee, E. Rios Velazquez, R. T. H. Leijenaar, C. Parmar, P. Grossmann, S. Carvalho, J. Bussink, R. Monshouwer, B. Haibe-Kains, D. Rietveld, F. Hoebers, M. M. Rietbergen, C. R. Leemans, A. Dekker, J. Quackenbush, R. J. Gillies, and P. Lambin, "Data from NSCLC-Radiomics," 2019. [Online]. Available: https://wiki.cancerimagingarchive.net/x/FgL1

[18] P. Jordan, P. M. Adamson, V. Bhattbhatt, S. Beriwal, S. Shen, O. Radermecker, S. Bose, L. S. Strain, M. Offe, D. Fraley, S. Principi, D. H. Ye, A. S. Wang, J. Heteren, N.-J. Vo, and T. G. Schmidt, "Pediatric chest-abdomen-pelvis and abdomen-pelvis CT images with expert organ contours," *Medical Physics*, Feb. 2022. [Online]. Available: https://doi.org/10.1002/mp.15485

[19] P. Jordan, P. M. Adamson, V. Bhattbhatt, S. Beriwal, S. Shen, O. Radermecker, S. Bose, L. S. Strain, M. Offe, D. Fraley, S. Principi, D. H. Ye, A. S. Wang, J. Van Heteren, N.-J. Vo, and T. G. Schmidt, "Pediatric chest/abdomen/pelvic CT exams with expert organ contours (Pediatric-CT-SEG)," 2021. [Online]. Available: https://wiki.cancerimagingarchive.net/x/jIFPBQ

[20] J. Yang, H. Veeraraghavan, S. G. Armato, K. Farahani, J. S. Kirby, J. Kalpathy-Kramer, W. van Elmpt, A. Dekker, X. Han, X. Feng, P. Aljabar, B. Oliveira, B. van der Heyden, L. Zamdborg, D. Lam, M. Gooding, and G. C. Sharp, "Autosegmentation for thoracic radiation treatment planning: A grand challenge at AAPM 2017," *Medical Physics*, vol. 45, no. 10, pp. 4568–4581, Sep. 2018. [Online]. Available: https://doi.org/10.1002/mp.13141

[21] J. Yang, G. Sharp, H. Veeraraghavan, W. Van Elmpt, A. Dekker, T. Lustberg, and M. Gooding, "Data from lung CT segmentation challenge," 2017. [Online]. Available: https://wiki.cancerimagingarchive.net/x/e41yAQ

[22] E. Brion, J. Léger, A. Barragán-Montero, N. Meert, J. A. Lee, and B. Macq, "Domain adversarial networks and intensity-based data augmentation for male pelvic organ segmentation in cone beam CT," *Computers in Biology and Medicine*, vol. 131, p. 104269, Apr. 2021. [Online]. Available: https://doi.org/10.1016/j.compbiomed.2021.104269

[23] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, no. 56, pp. 1929–1958, 2014. [Online]. Available: http://jmlr.org/papers/v15/srivastava14a.html

[24] L. Liu, J. Cheng, Q. Quan, F.-X. Wu, Y.-P. Wang, and J. Wang, "A survey on U-shaped networks in medical image segmentations," *Neurocomputing*, vol. 409, pp. 244–258, Oct. 2020. [Online]. Available: https://doi.org/10.1016/j.neucom.2020.05.070

[25] P. Singh, M. K. Singh, R. Singh, and N. Singh, "Federated learning: Challenges, methods, and future directions," in *Federated Learning for IoT Applications*. Springer International Publishing, 2022, pp. 199–214. [Online]. Available: https://doi.org/10.1007/978-3-030-85559-8_13