Collaborative Multi-Institutional Prostate Lesion Segmentation from MR images Using Deep Federated Learning Framework

Isaac Shiri, Eman Showkatian, Reza Mohammadi, Behrooz Razeghi, Soroush Bagheri, Ghasem Hajianfar, Yazdan Salimi, Mehdi Amini, Mostafa Ghelich Oghli, Sohrab Ferdowsi, Slava Voloshynovskiy, Habib Zaidi, *IEEE Fellow*

Abstract– **To develop a robust and generalizable deep learning (DL) model gathering a massive and heterogenous dataset is crucial as the DL performances could be varied across different acquisition and reconstruction settings in the real clinical situation. Furthermore, sharing data with third parties is highly limited because of legal, ethical, security, and privacy issues. To address the aforementioned challenges, Federated learning (FL) allows one to train a DL model without sharing the data between different centers in a distributed and decentralized manner. In the current study, we developed deep FL-based models for intraprostatic lesion segmentation using FL approaches and compared their results by center based. Altogether 400 histologically proven prostate cancer patients with T2-weighted MRI images from eight different centers were enrolled. Dynamic data augmentation techniques for flipping left/right, elastic deformation, and random cropping was adopted. Two-stage cascaded U-Net consisting of modified 3D U-Ne, and Dual Attention 2D U-Net were implemented as the core of DL segmentation. MRI images and a prostate mask were used as input in this network. In addition, Federated Averaging (FedAvg) algorithm was implemented in this study. All evaluations were performed on 30% of test sets. In terms of dice and Jaccard coefficient, CeBa and achieved 0.77 ± 0.06 vs. 0.84 ± 0.05 (CI95%: 0.76 - 0.79 vs. 0.82 - 0.85) and 0.63 ± 0.09 vs. 0.72 ± 0.08 and (CI95%:0.61 - 0.65 vs. 0.70 - 0.74). We set out to develop a DL-based automated algorithm capable of segmenting intra-prostatic lesions using T2W MR images. Due to the complex structure and low-contrast nature of the intraprostatic lesion in T2W MR images, our proposed algorithm performed very well compared to manual segmentation across different centers. Our FL algorithms outperformed center base algorithms in which each center developed a model using their local dataset, which addresses data sharing between different centers.**

Index Terms— **Federated learning, Deep learning, MRI, Prostate**

I. Shiri, Y. Salimi, M. Amini and H. Zaidi are with the Division of Nuclear Medicine and Molecular Imaging, Geneva University Hospital, 1211, Geneva 4, Switzerland (e-mail: isaac.shirilord@unige.ch, yazdan.salimi@unige.ch, mehdi.amini.1994.ma@gmail.com, habib.zaidi@hcuge.ch).

E. Showkatian and M. Ghelich Oghli are with Research and Development Department, Med Fanavarn Plus Co, Karaj, Iran (e-mail: i.shokatian@gmail.com, m.g31.mesu@gmail.com)

R.Mohammadi is with Department of Medical Physics, Iran University of Medical Sciences, Tehran, Iran (e-mail: reza021mohammadi@gmail.com,

B. Razeghi, S. Ferdowsi, D. Teodoro and S. Voloshynovskiy are with the University of Geneva, Switzerland (e-mail: behrooz.razeghi@unige.ch, s.ferdowsi@unige.ch, svolos@unige.ch)

S. Bagheri is with Department of Medical Physics, Kashan University of Medical Sciences, Kashan, Iran (e-mail: soroushb6975@gmail.com)

G. Hajianfar is with Rajaie Cardiovascular Medical and Research Center, Iran University of Medical Sciences, Tehran, Iran (e-mail: hajianfar.gh@gmail.com)

I. INTRODUCTION

Prostate cancer will be the third cause of cancer mortality in the United States in 2022. The estimated new cancers and deaths were 268,490 and 34,500, leading to 27% and 21% of all cancers in American men, respectively [1]. One of the main tools for diagnosing prostate cancer is a biopsy guided by transrectal ultrasound (TRUS) with fusion with magnetic resonance imaging (MRI) which is real-time, low-cost, and versatile. MRI offers increasingly reliable visualization of potentially significant prostate cancers [2]. Thus, it has shown advantages to better select patients for biopsy and facilitates direct targeting of lesions during the biopsy. The delineation of lesions manually is challenging due to inter and intra-observer variations[3-8].

To develop a robust and generalizable deep learning (DL) model gathering a massive and heterogenous dataset is crucial as the DL performances could be varied across different acquisition and reconstruction settings in the real clinical situation [3-6, 9-13]. However, single centers are unlikely able to provide these massive and heterogenous data because of limited resources and most likely gathering data with the same scanner, acquisition, and reconstruction setting [14-16]. Therefore, different centers and hospitals must collaborate and pool their dataset to third parties to gather this large and heterogenous dataset. However, sharing data with third parties is highly limited because of legal, ethical, security, and privacy issues [17].

FL brings this opportunity to train a DL model without sharing the data between different centers in a distributed and decentralized manner to address the aforementioned challenges. In the current study, we developed deep FL-based models for intraprostatic lesion segmentation using FL approaches and compared their results by center-based (training and evaluation performed in single centers).

II. MATERIALS AND METHODS

In this study, T2-weighted MRI images were taken. Altogether 400 histologically proven prostate cancer patients with T2-weighted MRI images from eight different centers were enrolled. All data were split in each center into a train/validation set (70/10% patients) and a test set (20% patients). Dynamic data augmentation techniques for flipping left/right, elastic deformation, and random cropping was adopted. Two-stage cascaded U-Net consisting of modified 3D U-Ne, and Dual Attention 2D U-Net were implemented as the core of deep learning segmentation. MRI images and a prostate mask were used as input in this network.
Federated Averaging (FedAvg) algorithm was implemented in this study. In FedAvg, the global model developed by the

server distributes data through different centers. Next, the models are trained separately in each center using the local data set, and finally, trained models from all centers are returned to the server to aggregate and update the central global model. These steps are repeated until convergence criteria are met, for example, until no significant loss descent is observed. In addition to FedAvg, we trained and tested models for each center separately, named center-based training (CeBa).

We optimized our loss function (binary cross-entropy + Dice-loss) by the Adaptive Moment Estimation (Adam) optimizer and used an initial learning rate of 5×10^{-4}, batch size of 2, and epoch value of 1000 with a linear warmup of 10 epochs. Furthermore, we implemented different standard quantitative segmentation metrics, including Dice Coefficient metric, Jaccard index, Sensitivity, Specificity, Precision, Average Surface Distance (ASD, mm), average Hausdorff Distance (avgHD, mm), 95% Hausdorff Distance (HD 95%, mm), Relative Absolute Volume Difference (RAVD, mm), and Volume Correlation to quantify the performance of DL models in different frameworks. All these metrics were calculated on test sets (20% of each center's data).

III. RESULTS

In terms of dice and Jaccard coefficient, CeBa and achieved 0.77 ± 0.06 vs. 0.84 ± 0.05 (CI95%: 0.76 - 0.79 vs. 0.82 - 0.85) and 0.63 ± 0.09 vs. 0.72 ± 0.08 and (CI95%:0.61 - 0.65 vs. 0.70 - 0.74). In terms of Sensitivity, Specificity, Precision, ASD, avgHD, HD (95%), RAVD, and VC 0.72 ± 0.09, 1 ± 0, 0.85 ± 0.11, 0.42 ± 0.17, 0.61 ± 0.31, 2.13 ± 1.59, -0.13 ± 0.21, and 0.97 ± 0.02 achieved for CeBa, respectively. And 0.8 ± 0.08, 1 ± 0, 0.89 ± 0.1, 0.3 ± 0.13, 0.41 ± 0.2, 1.59 ± 1.02, -0.08 ± 0.17, and 0.99 ± 0.01, respectively.

We depict an example of 3D rendered volumes of segmentation of lesion segmentation in Figure 1 for different frameworks of CeBa (yellow), FedAvg (purple), and Manual (red). Figure 2 represents 2D axial views of different patients (magnified to enter manual segmentation for better visualization). As depicted in these figures, the segmentation provided by different frameworks agrees with manual segmentation.
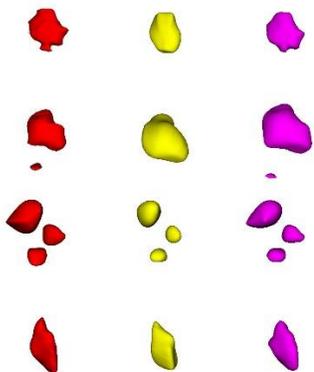
Fig 1. Example of 3D rendered volumes of prostate lesions segmentation. CeBa (yellow), FedAvg (purple), and Manual (red).
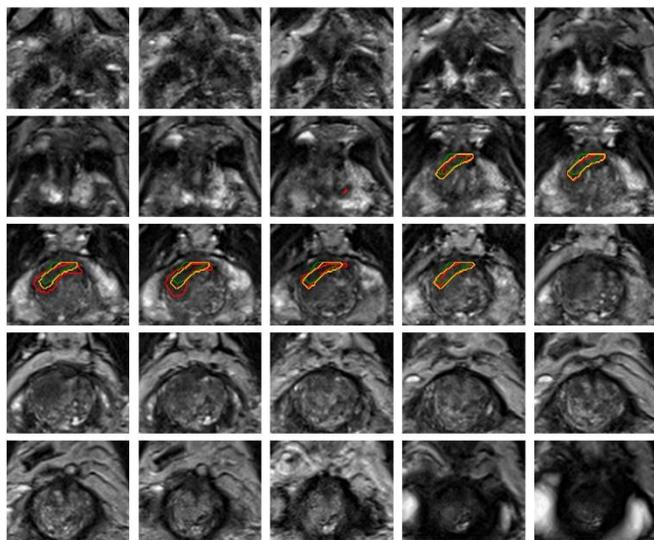
Fig 2. Example of 2D slices showing segmentation of prostate lesions. Red: manual segmentation, green: CeBa, and yellow: FedAvg

IV. DISCUSSION AND CONCLUSION

Large and heterogenous medial images from different centers using different scanners, acquisition, and reconstruction is essential for robust and generalizable DL modeling. Collecting this large and heterogenous data requires collaboration between different centers, which is limited because of the sensitive nature of the medical dataset and legal and ethical problems. FL allows us to develop DL models in a decentralized and distributed manner in which data of patients will keep in each center. We set out to develop a DL-based automated algorithm capable of segmenting intra-prostatic lesions using T2W MR images. Due to the complex structure and low-contrast nature of the intraprostatic lesion in T2W MR images, our proposed algorithm performed very well compared to manual segmentation across different centers.

REFERENCES

[1] R. L. Siegel, K. D. Miller, H. E. Fuchs, and A. Jemal, "Cancer statistics, 2022," (in eng), *CA Cancer J Clin,* vol. 72, no. 1, pp. 7-33, Jan 2022, doi: 10.3322/caac.21708.

[2] P. D. Q. A. T. E. Board, "Prostate Cancer Treatment (PDQ®): Patient Version," in *PDQ Cancer Information Summaries*. Bethesda (MD): National Cancer Institute (US), 2002.

[3] N. Sirjani *et al.*, "Automatic cardiac evaluations using a deep video object segmentation network," (in eng), *Insights Imaging,* vol. 13, no. 1, p. 69, Apr 8, 2022, doi: 10.1186/s13244-022-01212-9.

[4] V. Ashkani Chenarlogh *et al.*, "Clinical target segmentation using a novel deep neural network: double attention Res-U-Net," (in eng), *Sci Rep,* vol. 12, no. 1, p. 6717, Apr 25, 2022, doi: 10.1038/s41598-022-10429-z.

[5] V. Ashkani Chenarlogh *et al.*, "Fast and Accurate U-Net Model for Fetal Ultrasound Image Segmentation," (in eng), *Ultrason Imaging,* vol. 44, no. 1, pp. 25-38, Jan 2022, doi: 10.1177/01617346211069882.

[6] M. Ghelich Oghli *et al.*, "Automatic fetal biometry prediction using a novel deep convolutional network

architecture," (in eng), *Phys Med,* vol. 88, pp. 127-137, Aug 2021, doi: 10.1016/j.ejmp.2021.06.020.

[7]   I. Shiri, H. Arabi, A. Sanaat, E. Jenabi, M. Becker, and H. Zaidi, "Fully Automated Gross Tumor Volume Delineation From PET in Head and Neck Cancer Using Deep Learning Algorithms," (in eng), *Clin Nucl Med,* vol. 46, no. 11, pp. 872-883, Nov 1 2021, doi: 10.1097/rlu.0000000000003789.

[8]   M. A. Bjurlin *et al.*, "Update of the standard operating procedure on the use of multiparametric magnetic resonance imaging for the diagnosis, staging and management of prostate cancer," *The Journal of urology,* vol. 203, no. 4, pp. 706-712, 2020.

[9]   A. Sanaat, I. Shiri, S. Ferdowsi, H. Arabi, and H. Zaidi, "Robust-Deep: A Method for Increasing Brain Imaging Datasets to Improve Deep Learning Models' Performance and Robustness," (in eng), *J Digit Imaging,* vol. 35, no. 3, pp. 469-481, Jun 2022, doi: 10.1007/s10278-021-00536-0.

[10]   A. Akhavanallaf, R. Mohammadi, I. Shiri, Y. Salimi, H. Arabi, and H. Zaidi, "Personalized brachytherapy dose reconstruction using deep learning," (in eng), *Comput Biol Med,* vol. 136, p. 104755, Sep 2021, doi: 10.1016/j.compbiomed.2021.104755.

[11]   M. Salehi, A. Vafaei Sadr, S. R. Mahdavi, H. Arabi, I. Shiri, and R. Reiazi, "Deep Learning-based Non-rigid Image Registration for High-dose Rate Brachytherapy in Inter-fraction Cervical Cancer," (in eng), *J Digit Imaging,* Nov 23 2022, doi: 10.1007/s10278-022-00732-6.

[12]   H. Zaidi, I. Shiri, and N. Andratschke, "Two-Year Overall Survival Prediction in Non–Small-Cell Lung Cancer Patients Using Pre-Treatment Computed Tomography Images and Deep Neural Networks: A Multicentric Study," in *Medical Imaging with Deep Learning*, 2022.

[13]   Z. Yan, J. Wicaksana, Z. Wang, X. Yang, and K.-T. Cheng, "Variation-aware federated learning with multi-source decentralized medical image data," *IEEE Journal of Biomedical and Health Informatics,* vol. 25, no. 7, pp. 2615-2628, 2020.

[14]   A. Jabbarpour, S. R. Mahdavi, A. Vafaei Sadr, G. Esmaili, I. Shiri, and H. Zaidi, "Unsupervised pseudo CT generation using heterogenous multicentric CT/MR images and CycleGAN: Dosimetric assessment for 3D conformal radiotherapy," (in eng), *Comput Biol Med,* vol. 143, p. 105277, Jan 31 2022, doi: 10.1016/j.compbiomed.2022.105277.

[15]   I. Shiri *et al.*, "Decentralized Distributed Multi-institutional PET Image Segmentation Using a Federated Deep Learning Framework," (in eng), *Clin Nucl Med,* vol. 47, no. 7, pp. 606-617, Jul 1 2022, doi: 10.1097/rlu.0000000000004194.

[16]   I. Shiri *et al.*, "Decentralized collaborative multi-institutional PET attenuation and scatter correction using federated deep learning," *European Journal of Nuclear Medicine and Molecular Imaging,* vol. In press 2022, doi: https://doi.org/10.1007/s00259-022-06053-8.

[17]   M. Adnan, S. Kalra, J. C. Cresswell, G. W. Taylor, and H. R. Tizhoosh, "Federated learning and differential privacy for medical image analysis," *Scientific reports,* vol. 12, no. 1, pp. 1-10, 2022.