Advancing Lung Cancer Diagnosis: Federated Learning-Based Privacy Innovations

Zixiang Hao[a]
School of Science, Harbin Institute of Technology, Liuxian Avenue, Shenzhen, China

Keywords: Lung Cancer Treatment, Federated Learning, FL+NN Technique, Data Privacy.

Abstract: Lung cancer, as one of the most prevalent and lethal forms of cancer, presents a significant challenge to global healthcare systems. In recent years, the application of federated learning in lung cancer treatment has gained traction, offering several advantages. Federated learning addresses concerns regarding data privacy and security by allowing local model training on patient data, thereby minimizing the risk of privacy breaches. Furthermore, it facilitates the inclusion of diverse datasets from various healthcare institutions, enabling more comprehensive and representative model training. By analysing and summarizing the three methods—the Federated Learning (FL) + Neural Network (NN) technique (the FL+NN technique), the convolutional IT-2 fuzzy rough federated learning-neural architecture search model (the CIT2FR-FL-NAS model), and U-Net, the article underscores the potential of federated learning to revolutionize lung cancer therapy. The FL+NN technique combines federated learning with neural network models, demonstrating high accuracy in lung cancer classification. The CIT2FR-FL-NAS model integrates federated learning, neural architecture search, and fuzzy rough set theory to achieve accurate classification results while safeguarding privacy and reducing network complexity. Similarly, U-Net, a fully convolutional network architecture, shows effectiveness in segmenting organs in medical imaging, such as the heart and lungs. The potential is shown by the ability of enhancing accuracy, privacy, and collaboration in medical data analysis and treatment planning. The objective of the article is to stimulate further research and innovation in this critical healthcare domain.

1 INTRODUCTION

As one of the most prevalent and lethal forms of cancer, lung cancer poses a daunting threat to healthcare systems globally. Through traditional treatment methods, such as chemotherapy and radiotherapy, there have been significant strides in addressing lung cancer. However, these methods often come with many drawbacks, including high medical costs, adverse side effects, and inconsistent treatment outcomes, thereby prompting the exploration of alternative approaches.

Nowadays, various approaches such as artificial neural network have been applied to analyze large-scale patient datasets and develop personalized treatment strategies (Qiu, 2022). Despite notable advancements, traditional data analysis methods face challenges concerning data privacy, security, and interoperability across different healthcare institutions (Gupta et al., 2019). These limitations have spurred the exploration of innovative approaches that can harness the collective intelligence of distributed data sources without compromising patient privacy and data security.

In recent years, there has been growing interest in utilizing advanced technologies to enhance the effectiveness and efficiency of lung cancer treatment. One such technology is federated learning, which is a decentralized machine learning technique. It facilitates collaborative training of models among the multiple servers without the need to exchange sensitive patient data (Konečný et al., 2016). The shift towards data sharing and model training offers unprecedented opportunities for healthcare that is tailored to individuals and based on data analysis.

In the context of lung cancer treatment, federated learning offers several distinct advantages over traditional approaches. Firstly, it addresses concerns regarding data privacy and security by allowing models to be trained locally on patient data,

[a] https://orcid.org/0009-0004-3156-8333

Hao, Z.
Advancing Lung Cancer Diagnosis: Federated Learning-Based Privacy Innovations.
DOI: 10.5220/0012938800004508

In *Proceedings of the 1st International Conference on Engineering Management, Information Technology and Intelligence (EMITI 2024)*, pages 399-403
ISBN: 978-989-758-713-9

minimizing the risk of data breaches or privacy violations (McMahan et al., 2017). Additionally, federated learning facilitates the inclusion of diverse datasets from various healthcare institutions, thereby enabling more comprehensive and representative model training (Li et al., 2020). This aspect is particularly crucial in lung cancer treatment, where patient demographics, genetic profiles, and treatment responses can vary significantly. Furthermore, federated learning fosters collaborative research and knowledge sharing among healthcare providers and researchers, leading to accelerated innovation and improved treatment outcomes (Sheller et al., 2018). By pooling knowledge and expertise from multiple sources, federated learning enables the development of robust and generalizable models for lung cancer diagnosis, prognosis, and treatment planning. Moreover, the decentralized nature of federated learning ensures that the resulting models are adaptable to evolving patient needs and healthcare practices (Briggs et al., 2020).

In this review, the extensive application of federated learning in lung cancer treatment will be thoroughly explored. The objective is to delve into the fundamental principles of federated learning, assess existing methodologies and techniques, and analyze both the potential advantages and challenges associated with implementing this approach in lung cancer therapy. By shedding light on how federated learning can revolutionize lung cancer treatment, the review hopes to stimulate further research and innovation in this critical healthcare domain.

2 METHOD

2.1 Federated Learning Fundamentals

At the core of federated learning lies the principle of decentralized machine learning, wherein models are trained collaboratively across multiple devices or servers without the need for centralized data aggregation. This approach ensures data privacy and security by allowing model training to occur locally on individual devices or within separate healthcare institutions. The federated learning process typically involves several key steps.

2.1.1 Client Selection

Normally, a global model is initially created and distributed to participating devices or servers. These devices can be smartphones, tablets, or even edge computing nodes located within different healthcare institutions. Healthcare institutions or devices participating in federated learning are selected based on predefined criteria, such as data quality, patient population diversity, and computational capabilities (Li et al., 2020).

2.1.2 Local Model Training

Each selected client independently trains a local model using its own patient data while keeping the sensitive data securely stored on-device. The local model is updated iteratively through multiple epochs using standard machine learning algorithms, such as gradient descent.

2.1.3 Model Aggregation

After completing local model training, instead of transmitting raw data that may contain personally identifiable information, only the updated model parameters are securely transmitted to a centralized server or aggregator for aggregation. The server aggregates the model updates using techniques like Federated Averaging (FedAvg) or Federated Proximal (FedProx) to generate an improved global model that incorporates knowledge from all participating clients (McMahan et al., 2017).

2.1.4 Global Model Update

The central server then distributes this enhanced global model back to all participating clients for further iterations, normally a new round of local model training. This iterative process continues until convergence, or a predefined stopping criterion is met. During the process, all participating devices have collectively contributed their knowledge.

2.2 Models

2.2.1 Federated Learning-Based Method

The Federated Learning (FL) + Neural Network (NN) technique (the FL+NN technique), demonstrates promising performance in the classification of lung cancer. The use of deep learning techniques, such as NN models, enhances the performance of the FL+NN technique in lung cancer classification and diagnosis. The decentralized topology and distributed computing in the FL+NN approach facilitate faster and more secure computations, improving the overall performance of the technique. The approach achieves an accuracy of 89.63% in lung cancer classification, outperforming other models such as Support Vector Machine (SVM), K-Nearest Neighbour (KNN), and

Deep Neural Networks (DNN) in terms of accuracy, sensitivity, and specificity (Subashchandrabose et al., 2023). Among other models, DNN has the best performance compared with SVM and KNN. It even has a higher value than FL+NN on the computation accuracy of centralized server-based classification of lung cancer dataset. However, the FL+NN technique generally performs better. The FL+NN technique also ensures data privacy and security while utilizing distributed data, making it a reliable and efficient approach for lung cancer classification.

2.2.2 CIT2FR-FL-NAS-Based Method

Convolutional IT-2 fuzzy rough federated learning (CIT2FR-FL) is a framework that combines Convolutional Neural Networks (CNNs) with IT-2 fuzzy rough set theory in the context of federated learning (Liu et.al, 2022). Neural Architecture Search (NAS) is a technique utilized to automatically seek out optimal network architectures for deep learning models. Having been successfully applied in various domains, including image classification and medical data analysis, NAS can be performed using various methods including evolutionary algorithms and neuro-evolution (Jin et al., 2019).

The CIT2FR-FL-NAS model is a multi-objective convolutional IT-2 fuzzy rough federated learning framework with the goal of achieving high accuracy in medical data security while safeguarding privacy and reducing network complexity. The model employs a multi-objective evolutionary algorithm to automatically search for optimal network architectures for medical diagnostic problems. Each participant in the federated learning process trains the model locally using their own data, ensuring the privacy of patient information. Furthermore, the CIT2FR-FL-NAS model combines the interpretability of deep neural networks with the IT-2 fuzzy rough set theory, enhancing the interpretability of the convolutional neural network used for feature extraction from histopathological images. By integrating federated learning, neural architecture search, and fuzzy rough set theory, the CIT2FR-FL-NAS model achieves accurate classification results while reducing network complexity and protecting medical data security.

2.2.3 U-Net-Based Method

U-Net is a fully convolutional network architecture used for image segmentation in medical imaging (Siddique et al., 2021). It consists of a contracting path and an expanding path, forming a U-shape. Furthermore, it is trained using a pixel-wise binary cross-entropy loss function, comparing the predicted segmentation mask with the ground truth. Nowadays U-Net has been used for the segmentation of organs such as the heart and lungs in CT scan images. It has also been applied to the precise localization of organs at risk in radiotherapy, where accurate segmentation is crucial to avoid damaging side effects. The model is trained on large datasets, such as the non-small cell lung cancer-radiomics dataset (the NSCLC-Radiomics dataset), using federated learning to ensure privacy and security of patient data (Misonne et.al).

NSCLC-Radiomics dataset, which includes manual delineations of the gross tumor volume and segmentations of the lungs, heart, and esophagus for a subset of patients, contains 422 NSCLC patients. The performance of U-Net using the NSCLC-Radiomics dataset was evaluated using the Dice Similarity Coefficient (DSC3D). The results showed that the federated equal-chances variant of federated learning improved the segmentation performance on unbalanced datasets, achieving a DSC3D value of 0.879 for the heart segmentation. U-Net demonstrated its effectiveness in segmenting the heart using the NSCLC-Radiomics dataset, and the combination of U-Net with Federated Learning showed potential for improving medical image segmentation.

3 DISCUSSIONS

In general, there are several benefits of applying federated learning in medical treatment. It ensures the privacy and confidentiality of patient information, which is paramount in healthcare settings. Through allowing model training to occur locally on individual devices or within separate healthcare institutions, sensitive patient data remains secure and protected from potential breaches or privacy violations. Besides, federated learning enables the aggregation of knowledge from multiple institutions, leading to the creation of more accurate and robust models. By incorporating diverse datasets from various healthcare institutions, the resulting models are more comprehensive and representative. During the period, it could also foster collaboration among researchers and institutions, promoting the development of advanced diagnostic tools and providing personalized treatment strategies for lung cancer patients. Collaborative efforts in model development and validation contribute to the continuous improvement of healthcare practices, leading to better patient outcomes and advancements in the field of oncology.

However, it also comes with challenges. From the perspective of data, the data distribution among clients differs greatly, which makes it challenging to train a global model representative of all data sources. Federated learning must address issues related to data clutter, efficiency, and varying data standards across different sources to ensure high-quality training data. In terms of model parsability, the parsability for customers can set various parameters and security measures to strike a balance in efficiency, performance, and privacy which warrants further exploration. Communication efficiency is also a challenge, especially with many clients who require effective communication protocols. In the training process of federated learning, frequent data transmission between the server and multiple clients, along with data encryption and decryption, consumes substantial communication bandwidth, potentially leading to transmission delays. Some more advanced hardware or transmission technologies should be considered (Deng, 2019; Sugaya, 2019). Given that federated learning aims to improve the performance of machine learning models by leveraging diverse datasets, ensuring model accuracy and precision across different data sources is a challenge that needs to be addressed. Besides, providing incentives for client devices to participate in federated learning tasks is crucial for the success of the process. Designing efficient incentive mechanisms can encourage data sharing while addressing self-interest concerns. There is also feasibility for the involvement of blockchain. The decentralized nature of blockchain enhances transparency and trust in data storage and processing, reducing the control of data by single entities. The integration with federated learning facilitates cross-organizational model training and sharing, enhancing model credibility and reliability. By combining blockchain's consensus mechanism with federated learning's model aggregation process, the computational burden of the federated learning system is notably reduced, ensuring an optimal solution for model aggregation.

4 CONCLUSIONS

Federated learning provides a promising approach to revolutionize lung cancer therapy by addressing data privacy, model accuracy, and collaboration challenges. It allows local model training on patient data, thus minimizes the risk of privacy breaches while enabling the inclusion of diverse datasets from various healthcare institutions. Through methods like the FL+NN technique, CIT2FR-FL-NAS model, and U-Net, federated learning demonstrates its potential in achieving accurate classification results while safeguarding patient privacy. Collaborative research and knowledge among healthcare stakeholders is enhanced, accelerating innovation in personalized treatment strategies. However, challenges such as data distribution disparities, communication efficiency, and incentivizing client participation remain. Therefore, there exists the necessity of further exploration and innovation. The integration of federated learning with other techniques such as blockchain offers opportunities to improve transparency and computational efficiency in model aggregation. Federated learning holds promise in improving patient outcomes and advancing oncology research, stimulating further exploration and innovation in this critical healthcare domain.

REFERENCES

Briggs, C., Wells, J., & Sharma, A. 2020. A Federated Learning Approach for Automated Lung Cancer Detection and Prediction. arXiv preprint arXiv: 2010.11565.

Deng, X., et al. 2019. Continuously frequency-tuneable plasmonic structures for terahertz bio-sensing and spectroscopy. Scientific reports, 9(1), 3498.

Jin, H., Song, Q., & Hu, X. 2019. Auto-keras: An efficient neural architecture search system. In Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining, 1946-1956.

Konečný, J., McMahan, H. B., Ramage, D., & Richtárik, P. 2016. Federated optimization: Distributed optimization beyond the datacenter. arXiv preprint arXiv:15 11.03575.

Li, T., Sahu, A. K., Zaheer, M., Sanjabi, M., Talwalkar, A., & Smith, V. 2020. Federated optimization in heterogeneous networks. arXiv preprint arXiv: 1812.06127.

Liu, X., et al. 2022. Federated neural architecture search for medical data security. IEEE transactions on industrial informatics, 18(8), 5628-5636.

McMahan, H. B., Moore, E., Ramage, D., Hampson, S., & y Arcas, B. A. 2017. Communication-efficient learning of deep networks from decentralized data. In Artificial Intelligence and Statistics, 1273-1282.

Misonne, T., & Jodogne, S. 2022. Federated Learning for organ segmentation. dial.uclouvain.be

Qiu, Y., Wang, J., Jin, Z., Chen, H., Zhang, M., & Guo, L. 2022. Pose-guided matching based on deep learning for assessing quality of action on rehabilitation training. Biomedical Signal Processing and Control, 72, 103323.

Sheller, M. J., Reina, G. A., Edwards, B., Martin, J., Bakas, S., & Kovacs, T. 2018. Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data. Scientific reports, 9(1), 1-12.

Siddique, N., et al. 2021. U-net and its variants for medical image segmentation: A review of theory and applications. IEEE Access, 9, 82031-82057.

Subashchandrabose, U., et al. 2023. Ensemble Federated learning approach for diagnostics of multi-order lung cancer. Diagnostics, 13(19), 3053.

Sugaya, T., Deng, X., 2019. Resonant frequency tuning of terahertz plasmonic structures based on solid immersion method. 2019 44th International Conference on Infrared, Millimeter, and Terahertz Waves, 1-2.