Equitable Federated Learning with NCA

Nick Lemke[1,⋆], Mirko Konstantin[⋆], Henry John Krumb, John Kalkhof,
Jonathan Stieber, Anirban Mukhopadhyay

Technical University of Darmstadt, Darmstadt, Germany
[1]Corresponding author: nick.lemke@tu-darmstadt.de

Abstract. Federated Learning (FL) is enabling collaborative model training across institutions without sharing sensitive patient data. This approach is particularly valuable in low- and middle-income countries (LMICs), where access to trained medical professionals is limited. However, FL adoption in LMICs faces significant barriers, including limited high-performance computing resources and unreliable internet connectivity. To address these challenges, we introduce FedNCA, a novel FL system tailored for medical image segmentation tasks. FedNCA leverages the lightweight Med-NCA architecture, enabling **training on low-cost edge devices**, such as widely available smartphones, while **minimizing communication costs**. Additionally, our encryption-ready FedNCA proves to be **suitable for compromised network communication**. By overcoming infrastructural and security challenges, FedNCA paves the way for inclusive, efficient, lightweight, and encryption-ready medical imaging solutions, fostering equitable healthcare advancements in resource-constrained regions. We make our implementation publicly available at: https://github.com/MECLabTUDA/FedNCA

Keywords: Federated Learning · Equity · Resource Limited

1 Introduction

Federated Learning (FL) is rapidly emerging as a transformative approach in medical imaging [1]. Unlike traditional machine learning methods that rely on centralized datasets, FL facilitates the collaborative training of models across multiple institutions without the need to share sensitive patient data [1]. This makes it particularly appealing for applications in healthcare, where adherence to stringent data privacy guidelines, such as HIPAA and GDPR, is paramount [2]. Even in low-and-middle-income countries (LMICs) with limited resources and a shortage of healthcare professionals, FL has the potential to provide access to high-quality medical AI [3].

While the benefits are evident, the adoption of FL in LMICs is hindered by several infrastructural and technical challenges [4]. One of the primary barriers in LMICs is the limited access to high-performance computing resources [5], making it difficult to train deep neural networks and implement traditional FL

[⋆] These authors contributed equally to this work.

frameworks. Additionally, slow and unreliable internet connections exacerbate the problem, as the exchange of model updates between clients and servers becomes time-intensive and inefficient [6].

In state-of-the-art FL solutions, models are becoming increasingly larger, with architectures growing more complex [7]. The rising computational demands necessitate powerful hardware, stable network infrastructures, and substantial energy resources — requirements that are often inaccessible in resource-constrained regions [8]. As state-of-the-art FL methods evolve, the risk of exacerbating global disparities in AI accessibility increases, underscoring the urgency for lightweight, adaptable, and efficient FL frameworks tailored to LMICs [9].

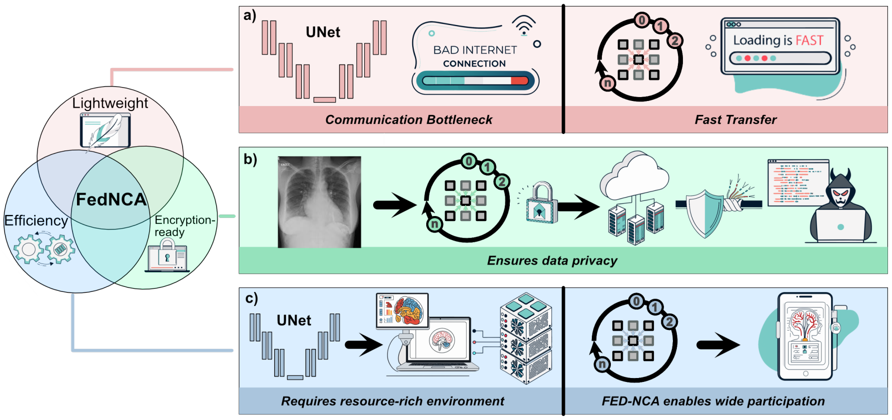

Fig. 1: FedNCA a) circumvents problems with low bandwidth internet connections, b) allows for efficient encryption to protect from adversaries or untrusted servers, c) is trainable on diverse hardware including smartphones.

To address these challenges, we propose FedNCA, an **efficient**, **lightweight**, and **encryption-ready** FL system for LMICs (Fig. 1). Unlike conventional FL systems that require substantial computational resources and high-bandwidth connectivity, FedNCA is optimized for deployment in resource-constrained environments. At its core, FedNCA leverages the Med-NCA backbone [10], which is inherently lightweight and efficient. This approach allows training on low-cost edge devices, such as inexpensive smartphones [11], which are available even in resource-constrained regions [12]. Moreover, the NCA architecture has 5000× fewer parameters than a UNet, thereby lowering communication costs and enhancing accessibility in regions with limited internet bandwidth. This not only significantly cuts operational costs but also democratizes access to cutting-edge medical AI, allowing broader participation in FL without the barriers of expensive hardware and network constraints.

Another crucial challenge is the presence of untrusted or even malicious servers [13], particularly in rural and remote regions where none of the trusted peers can act as servers. In such scenarios, ensuring data privacy and security becomes a significant concern. Encryption techniques like homomorphic encryption provide strong privacy protection guarantees, even against untrusted or malicious servers [14]. However, despite its robust security benefits, homomorphic encryption introduces substantial computational overhead [15], which can vary depending on the size of the message, potentially limiting its practicality in resource-constrained environments. Due to the lightweight MedNCA architecture, the *time required for the encryption and decryption process is reduced by a factor of 1800* compared to state-of-the-art architectures like TransUnet [16]. By drastically reducing computational overhead, FedNCA provides strong privacy guarantees, making secure FL more accessible to LMICs. Consequently, **robust privacy protection is no longer exclusive to those with high computational resources** but becomes available in regions that lack powerful infrastructure.

To address both, the infrastructure and security challenges associated with FL in LMICs we make the following contributions in this work: 1) We propose a secure and communication-efficient FL algorithm specifically for NCAs, 2) We evaluate the trained models in terms of their segmentation performance and the transmission costs coming up during FL, 3) We study the real-world applicability on smartphones, and 4) we analyze the efficiency improvement of the homomorphic encryption on NCAs.

2 Background

Neural Cellular Automata are characterized by their minimal number of parameters, making them highly efficient compared to traditional neural network architectures. Med-NCA [10], an NCA-based architecture specifically designed for medical image segmentation, has demonstrated performance comparable to conventional neural networks, like the UNet [17] while utilizing only a small fraction of the parameters. Med-NCA's efficiency stems from its iterative application of simple rules. Cellular automata like Conway's Game of Life are known to be Turing-complete, demonstrating that complex computations can be achieved with inherently simple update rules. NCA-based algorithms adapt this idea and encode the rule in a lightweight neural architecture [18], which significantly reduces hardware requirements compared to large architectures.

Communication Bottlenecks in FL arise due to its star-shaped topology, where a central server receives updates from multiple clients and then sends aggregated updates back [19]. These updates typically consist of model parameters or gradients, which are often large in size, leading to significant bandwidth consumption and increased latency [6]. This challenge is especially pronounced in settings with limited network capacity [20]. To mitigate this issue, various encoding and compression techniques have been introduced, such as quantization, sparsification, and federated dropout, which aim to reduce the size of updates while maintaining acceptable model performance [21]. However, these methods come with an inherent trade-off: while they decrease communication costs, they may also introduce a performance drop due to the loss of information in the transmitted updates [22].

3 Methodology

The deployment of secure FL systems in LMICs is often hindered by high communication overhead and computational demands. FedNCA addresses these challenges by providing an **efficient**, **lightweight**, and **encryption-ready** solution. These key attributes make FedNCA an ideal solution for resource-constrained settings. An outline of our FL algorithm can be seen in Fig. 2 and Alg. 1.

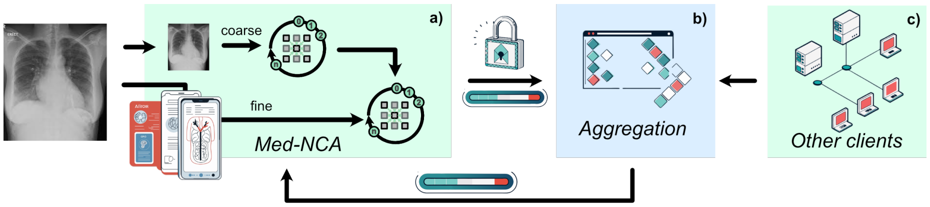

Fig. 2: Our FedNCA setup including a) the two-stage Med-NCA backbone, b) the aggregation of encrypted weights, and c) other clients connected via a weak internet connection.

Efficient: Inspired by Med-NCA, each client trains two NCAs via end-to-end backpropagation through time (BPTT). In detail, the downsampled (coarse) image is given to the first stage, which distributes knowledge on a global scale by running the first NCA f_θ for $T_0 = 20$ steps. After that, the hidden channels of the NCA are upscaled and concatenated with the high-resolution (fine) image. The second stage uses the global knowledge from the previous NCA to refine the segmentation via a second NCA f_ω in additional $T_1 = 40$ steps. During the forward pass, the deep learning engine automatically unrolls the computation of both NCAs f_θ and f_ω along time (steps) to create a computational graph. After computing the cross-entropy loss $\mathscr{L}$, the gradients are propagated back through time, averaging the gradients of each weight along each time step. Although BPTT is considered to be slow and inefficient, BPTT for NCAs is the opposite. As our NCAs are defined by inherently simple functions f_θ and f_ω, even low-energy devices like smartphones and tablets can train NCAs.

Lightweight: To further enhance communication efficiency, FedNCA's lightweight architecture ensures that clients transmit only small weight updates to the central server. With its compact 284KB model size, FedNCA drastically reduces bandwidth requirements, making it an ideal fit for FL in low connectivity environments (Fig. 2 c). This reduces communication costs by nearly 500× compared to traditional U-Net models, mitigating common FL communication bottlenecks.

Algorithm 1 FedNCA client and server functions.

```
– $\varphi$: Encryption function with corresponding key
– $f_\theta$, $f_\omega$: Backbone NCAs with weights $\theta, \omega$ and encrypted weights $\Theta, \Omega$
– $\mathscr{L}$: Segmentation loss function, $\eta$: Learning rate
– $x$, $y$: Image $x$ and segmentation target $y$ (for simplicity here only one)
– $T_0, T_1$: Number of steps for fine and course NCA
ClientUpdate($x$, $y$, $\{\Omega, \Theta\}$):
    $\{\theta, \omega\} \leftarrow \varphi^{-1}(\{\Theta, \Omega\}, \text{key})$        ▷ Decrypt parameters
    $z \leftarrow$ downscale($x$)
    for each step $s = 0...T_0$ do
        $z \leftarrow f_\theta(z)$
    end for
    $z \leftarrow$upscale($z, x$)
    for each step $s = 0...T_1$ do
        $z \leftarrow f_\omega(z)$
    end for
    $\theta \leftarrow \theta - \eta \nabla_\theta \mathscr{L}(x, y)$        ▷ Compute loss and update parameters
    $\omega \leftarrow \omega - \eta \nabla_\omega \mathscr{L}(x, y)$
    Return $\varphi(\{\theta, \omega\}, \text{key})$ to the server        ▷ Encrypt and return parameters
ServerUpdate($\{\Omega_0, \Theta_0\}$, ..., $\{\Omega_n, \Theta_n\}$):        ▷ Aggregate parameters from $n$ clients
    $\Theta \leftarrow \frac{1}{n}(\Theta_0 + ... + \Theta_n)$        ▷ Average encrypted parameters
    $\Omega \leftarrow \frac{1}{n}(\Omega_0 + ... + \Omega_n)$
    Return $\{\Theta, \Omega\}$ to the clients
```

This enables frequent updates even in low-connectivity environments, making it highly suitable for deployment in regions with limited internet access.

Encryption-ready: Homomorphic Encryption (HE) offers a quantum-proof level of security, ensuring privacy preservation even against an untrusted server. This is particularly crucial in FL where reconstruction [23] or source inference attacks [24] on client updates may leak private information about the client's training data. HE enables secure aggregation by allowing encryption φ and decryption φ^{-1} to behave as homomorphisms between plaintext and ciphertext, thereby satisfying

$$\varphi(m_1) * \varphi(m_2) = \varphi(m_1 * m_2) \qquad \forall m_1, m_2 \in M \tag{1}$$

where M represents all possible messages and $* \in \{\cdot, +\}$ represents a group operation, such as addition or multiplication.

As a result, the server can aggregate encrypted client updates without decrypting them, effectively eliminating the threat of server-side data leakage attacks. However, a major downside of HE is its high computational cost, which makes it impractical for large-scale client updates. To mitigate this, FedNCA's low number of parameters facilitates seamless homomorphic encryption, making FedNCA encryption-ready. In our setup, we utilized the CKKS [25] scheme, which is specifically designed for floating-point numbers, making it highly suitable for FL applications.

By combining efficiency, lightweight design, and encryption-readiness, FedNCA provides a scalable and secure FL solution for LMICs, addressing key barriers to AI adoption in resource-limited settings.

4 Experiments

Ultrasound: The *Fetal Abdominal Structures Segmentation* [26] dataset includes nearly 1500 images of fetal abdomen circumference (AC). Ultrasound images were captured following a standardized protocol, using Siemens Acuson, Voluson 730 (GE Healthcare Ultrasound), or Philips-EPIQ Elite (Philips Healthcare Ultrasound) systems. For our experiments, we focused on segmenting liver structures in the images. To create a challenging scenario with limited data per client, we reserve a random set of 118 (70%) patients for testing. The remaining 51 patients are randomly split among 5 FL clients.

XRay: The *MIMIC-III* [27] dataset consists of chest XRay images of patients in tertiary care. The segmentation encompasses both the left and right lungs. In our experiments, we utilize a random subset of 50 images, reserving 25 images for the test set and distributing the other 25 evenly among 5 clients.

Baselines: We compare FedNCA to federated UNet [17] and TransUNets [16], indicated by *Fed UNet*, and *Fed TransUNet*. Additionally, we use quantization and sparsification methods to reduce the transmission cost in the federated protocol. Specifically, we quantize the model weights to floating point values with 4-bit precision, indicated by *4-bit*. All other weights are encoded in 32-bit precision. Furthermore, we sparsify the model weights sent to the server by an unidirectional top-k algorithm [28]. Our algorithm selects only the most important client parameters, discarding the others. We consider parameters to be important by their difference to the values they had in the last FL round. We select the largest k% of parameters to be sent to the server in the upstream. The server, on the other hand, sends all aggregated parameters to the clients. We indicate this method by *top-k*, where k% indicates the number of parameters sent in each upstream.

Quantitative Scores: We measure the segmentation precision of each method using the *Dice* score. The *transmission cost* measures the amount of data (in MiB) transferred in each FL round, including the model parameters and the associated metadata introduced by the quantization or sparsification algorithms.

5 Results

In this section, we present our results when training in low-bandwidth regions. We investigate runtime when training on affordable hardware, and we measure runtime for encrypting weights of different segmentation models.

Training in low-bandwidth regions: The quantitative results presented in Fig. 3 provide insights into the Dice score and transmission costs of the model parameters. The results demonstrate that FedNCA consistently achieves equal segmentation quality of 74% and 78% to the baseline models, despite its 2000×

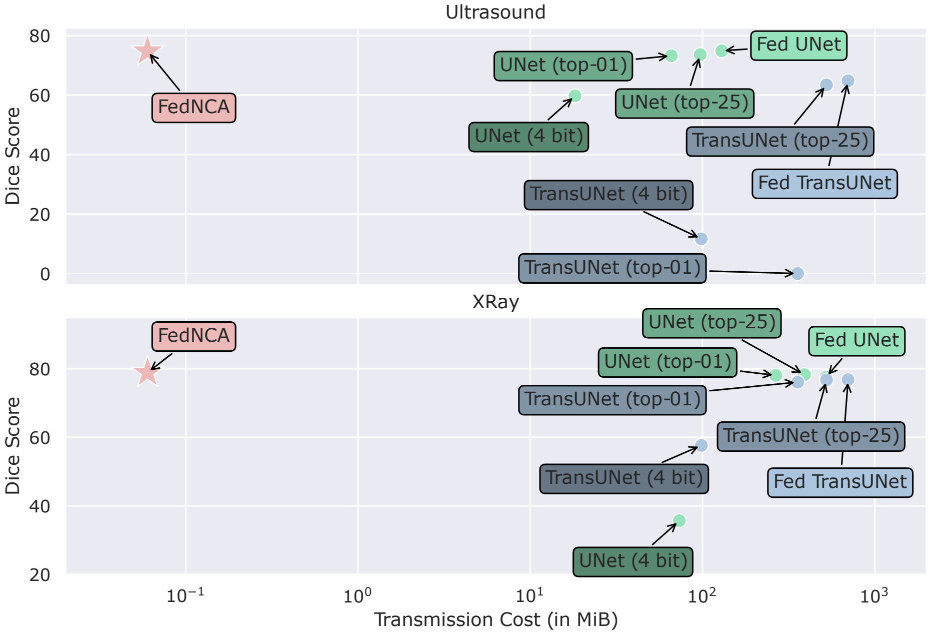

Fig. 3: Dice score and transmission cost in MiB. FedNCA achieves the best Dice while requiring the least transmission cost.

lower communication overhead. To address these high transmission costs, we apply compression techniques to the U-Net-based approaches. However, *even after compression, the transmission costs remain at least* 300× *higher* than those of FedNCA, indicating that our method is inherently more efficient without requiring additional compression. Furthermore, the compressed model updates experience significant performance degradation, revealing a clear tradeoff between reducing communication costs and maintaining model accuracy. FedNCA, on the other hand, is much more efficient without any decline in segmentation accuracy.

Training time on inexpensive hardware: We perform a case study to investigate the trainability of FedNCA on affordable devices. We report the time taken for training a single epoch on heterogeneous hardware. The results in Fig. 4 show that training on smartphones and tablets cheaper than 300€ is feasible.

Efficiency of homomorphic encryption: To preserve privacy during federated training, even in the case of an untrusted server, we encrypt the parameters using a homomorphic encryption scheme. In Fig. 5 we report a runtime analysis of the homomorphic encryption and decryption on FedNCA, UNet, and TransUNet on a recent Intel CPU. While the encryption scheme adds less than 20 milliseconds for FedNCA, encryption and decryption of the UNet is 1400× *slower*, adding 27 seconds to each FL round. For the bigger TransUNet, the runtime is 1800× *longer*, rendering FL slow and inefficient, especially for low-cost computing machinery.

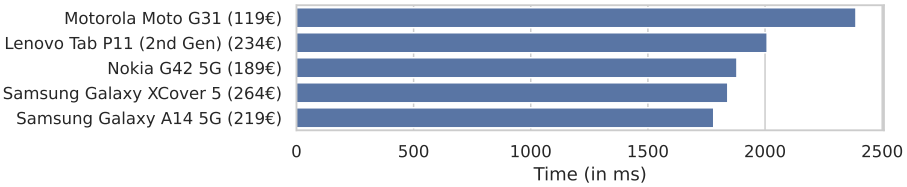

Fig. 4: Training time per epoch of FedNCA on heterogeneous hardware.

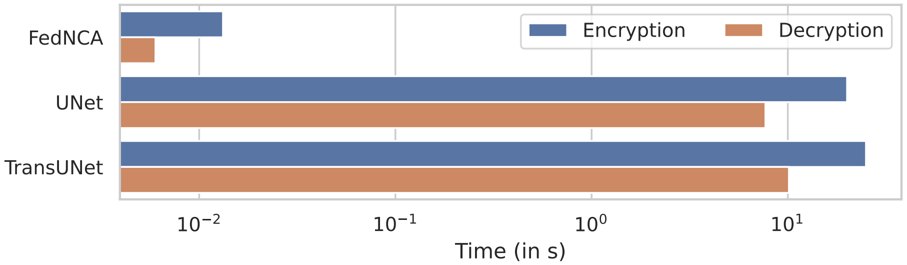

Fig. 5: Time (in s) taken for homomorphic encryption and decryption of FedNCA, UNet, and TransUNet on an Intel i7-13700K CPU.

6 Conclusion

In this work, we introduced FedNCA, a FL framework that leverages Neural Cellular Automata-based architecture for **efficient**, **lightweight**, and **encryption-ready** model training. Our experiments demonstrated robust performance across two segmentation tasks while showcasing efficient communication, enabled by the lightweight model architecture. Additionally, the low parameter count allowed the models to be trained on inexpensive smartphones, significantly reducing the participation burden within this framework. Furthermore, the reduced number of parameters enhances the efficiency of Homomorphic Encryption, making our solution encryption-ready and well-suited for secure, privacy-preserving applications. Traditional model architectures, such as U-Net and TransUNet, utilize compression methods to alleviate the communication bottleneck in FL, making it more accessible in regions with limited internet bandwidth. However, these compression techniques come with a tradeoff in performance. In contrast, FedNCA has been shown to be more communication-efficient while outperforming compressed versions of these models. The challenge of using encryption in traditional solutions arises from the high computational cost of encryption and decryption. Our results demonstrate that, due to its model architecture, FedNCA significantly reduces this computational burden, making it encryption-ready for real-world applications. FedNCA allows *equitable participation in AI training*, enabling even low-resource clinics and underrepresented populations to contribute without the requirement of data sharing. The tiny architecture runs on widely available devices, including smartphones, making FL accessible where traditional models are impractical. By combining efficient segmentation with federated training, FedNCA ensures AI models benefit from diverse global data, making high-quality medical AI available to all, regardless of infrastructure or location.

Acknowledgments. This work has been partially funded by the Federal Ministry of Education and Research project ”FED-PATH” (grant 01KD2210B) and by the hessian.AI project "Fed-NCA".

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Hao Guan, Pew-Thian Yap, Andrea Bozoki, and Mingxia Liu. Federated learning for medical image analysis: A survey. *Pattern Recognition*, page 110424, 2024.
2. Fatemeh Mosaiyebzadeh, Seyedamin Pouriyeh, Reza M Parizi, Quan Z Sheng, Meng Han, Liang Zhao, Giovanna Sannino, Caetano Mazzoni Ranieri, Jó Ueyama, and Daniel Macêdo Batista. Privacy-enhancing technologies in federated learning for the internet of healthcare things: a survey. *Electronics*, 12(12):2703, 2023.
3. Dinh C Nguyen, Quoc-Viet Pham, Pubudu N Pathirana, Ming Ding, Aruna Seneviratne, Zihuai Lin, Octavia Dobre, and Won-Joo Hwang. Federated learning for smart healthcare: A survey. *ACM Computing Surveys (Csur)*, 55(3):1–37, 2022.
4. Tadeusz Ciecierski-Holmes, Ritvij Singh, Miriam Axt, Stephan Brenner, and Sandra Barteit. Artificial intelligence for strengthening healthcare systems in low- and middle-income countries: a systematic scoping review. *npj Digital Medicine*, 5(1):162, 2022.
5. Raghavendra Selvan, Bob Pepin, Christian Igel, Gabrielle Samuel, and Erik B Dam. Equity through access: A case for small-scale deep learning. *arXiv preprint arXiv:2403.12562*, 2024.
6. Luke Melas-Kyriazi and Franklyn Wang. Intrinisic gradient compression for federated learning. *arXiv preprint arXiv:2112.02656*, 2021.
7. Yuxi Liu, Guibo Luo, and Yuesheng Zhu. FedFMS: Exploring Federated Foundation Models for Medical Image Segmentation . In *proceedings of Medical Image Computing and Computer Assisted Intervention – MICCAI 2024*, volume LNCS 15008. Springer Nature Switzerland, October 2024.
8. Chaoyang He, Alay Dilipbhai Shah, Zhenheng Tang, Di Fan1Adarshan Naiynar Sivashunmugam, Keerti Bhogaraju, Mita Shimpi, Li Shen, Xiaowen Chu, Mahdi Soltanolkotabi, and Salman Avestimehr. Fedcv: a federated learning framework for diverse computer vision tasks. *arXiv preprint arXiv:2111.11066*, 2021.
9. Ruixuan Liu, Fangzhao Wu, Chuhan Wu, Yanlin Wang, Lingjuan Lyu, Hong Chen, and Xing Xie. No one left behind: Inclusive federated learning over heterogeneous devices. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 3398–3406, 2022.
10. John Kalkhof, Camila González, and Anirban Mukhopadhyay. Med-nca: Robust and lightweight segmentation with neural cellular automata. In *International Conference on Information Processing in Medical Imaging*, pages 705–716. Springer, 2023.

11. John Kalkhof, Amin Ranem, and Anirban Mukhopadhyay. Unsupervised training of neural cellular automata on edge devices. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 498–507. Springer, 2024.
12. Mahmoud Elkasabi and Azam Khan. The evolution of mobile phone surveys in low- and middle-income countries: A study of coverage structure. *International Journal of Public Opinion Research*, 35(4):edad031, 10 2023.
13. Depeng Chen, Xiao Jiang, Hong Zhong, and Jie Cui. Building trusted federated learning: Key technologies and challenges. *Journal of Sensor and Actuator Networks*, 12(1), 2023.
14. Truc Nguyen and My T Thai. Preserving privacy and security in federated learning. *IEEE/ACM Transactions on Networking*, 32(1):833–843, 2023.
15. Weizhao Jin, Yuhang Yao, Shanshan Han, Jiajun Gu, Carlee Joe-Wong, Srivatsan Ravi, Salman Avestimehr, and Chaoyang He. Fedml-he: An efficient homomorphic-encryption-based privacy-preserving federated learning system. *arXiv preprint arXiv:2303.10837*, 2023.
16. Jieneng Chen, Yongyi Lu, Qihang Yu, Xiangde Luo, Ehsan Adeli, Yan Wang, Le Lu, Alan L Yuille, and Yuyin Zhou. Transunet: Transformers make strong encoders for medical image segmentation. *arXiv preprint arXiv:2102.04306*, 2021.
17. Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, pages 234–241. Springer, 2015.
18. Alexander Mordvintsev, Ettore Randazzo, Eyvind Niklasson, and Michael Levin. Growing neural cellular automata. *Distill*, 5(2):e23, 2020.
19. Jiajun Wu, Fan Dong, Henry Leung, Zhuangdi Zhu, Jiayu Zhou, and Steve Drew. Topology-aware federated learning in edge computing: A comprehensive survey. *ACM Computing Surveys*, 56(10):1–41, 2024.
20. Jaewon Yun, Yongjeong Oh, Yo-Seb Jeon, and H Vincent Poor. Communication-efficient federated learning over capacity-limited wireless networks. *IEEE Transactions on Cognitive Communications and Networking*, 2024.
21. Dingzhu Wen, Ki-Jun Jeon, and Kaibin Huang. Federated dropout—a simple approach for enabling federated learning on resource constrained devices. *IEEE wireless communications letters*, 11(5):923–927, 2022.
22. Grant Wilkins, Sheng Di, Jon C Calhoun, Zilinghan Li, Kibaek Kim, Robert Underwood, Richard Mortier, and Franck Cappello. Fedsz: Leveraging error-bounded lossy compression for federated learning communications. In *2024 IEEE 44th International Conference on Distributed Computing Systems (ICDCS)*, pages 577–588. IEEE, 2024.
23. Jonas Geiping, Hartmut Bauermeister, Hannah Dröge, and Michael Moeller. Inverting gradients-how easy is it to break privacy in federated learning? *Advances in neural information processing systems*, 33:16937–16947, 2020.
24. Hongsheng Hu, Zoran Salcic, Lichao Sun, Gillian Dobbie, and Xuyun Zhang. Source inference attacks in federated learning. In *2021 IEEE International Conference on Data Mining (ICDM)*, pages 1102–1107. IEEE, 2021.
25. Jung Hee Cheon, Andrey Kim, Miran Kim, and Yongsoo Song. Homomorphic encryption for arithmetic of approximate numbers. In *Advances in Cryptology–ASIACRYPT 2017: 23rd International Conference on the Theory and Applications of Cryptology and Information Security, Hong Kong, China, December 3-7, 2017, Proceedings, Part I 23*, pages 409–437. Springer, 2017.

26. Roberto; Santos Luís Otávio; Soares Muylaert Barroso Felipe; Zimmermann Loureiro Chaves Thiago Da Correggio, Karine Souza; Noya Galluzzo. Fetal abdominal structures segmentation dataset using ultrasonic images. *Mendeley Data*, 2023.
27. Alistair EW Johnson, Tom J Pollard, Lu Shen, Li-wei H Lehman, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G Mark. Mimic-iii, a freely accessible critical care database. *Scientific data*, 3(1):1–9, 2016.
28. Pengchao Han, Shiqiang Wang, and Kin K Leung. Adaptive gradient sparsification for efficient federated learning: An online learning approach. In *2020 IEEE 40th international conference on distributed computing systems (ICDCS)*, pages 300–310. IEEE, 2020.