SECURE IMAGE SEGMENTATION WITH FEDERATED LEARNING THROUGH KNOWLEDGE DISTILLATION

K.R. Nandhashree[1], J. Kalesh[2], U. Jashwanth[3], and S.R. Sai Arjun[4]
[1]Department of Cyber Security/SRM Valliammai Engineering College,
Chennai, India
Email: nandhashreekrcys@srmvalliammai.ac.in
[2-4]Department of Cyber Security/SRM Valliammai Engineering College,
Chennai, India
Email: {jkalesh20@gmail.com, jashwanth792003@gmail.com, saiarjun0619@gmail.com}

***Abstract*—Image segmentation is crucial in fields such as medical diagnostics, autonomous navigation, and satellite monitoring, however, centralized training approaches violate privacy as sensitive data must be shared. This work proposes an image segmentation framework that enables secure and efficient segmentations through a combination of Federated Learning (FL) and Knowledge Distillation (KD). Importantly, while FL maintains privacy by keeping the data on local devices, KD decreases computation and communication demands by enabling smaller models to learn from larger ones. Ultimately, our framework can achieve better segmentation performance while protecting the data in resource-constrained and heterogeneous environments. This mechanism allows for minimal data transmission and computation costs, which is a great advantage for usage in resource-constrained environments such as mobile and edge devices. The innovation in this project stems from the integration of Federated Learning and KD to create a powerful and privacy-preserving image segmentation framework, maintaining high performance standards in heterogeneous and constrained environments. The result is a scalable and secure solution for sensitive image segmentation tasks that can be applied to areas such as healthcare, autonomous vehicles, and remote sensing.**

***Index Terms*—Federated Learning, Knowledge Distillation, Image Segmentation, Privacy, Security, Distributed Computing, Edge Devices.**

I. INTRODUCTION

Image segmentation is an important aspect in various fields such as healthcare, autonomous driving, and remote sensing, as it allows machines to comprehend and process images accurately. over the years image segmentation models have been made to rely on centralized data processors thus raising considerable privacy concerns particularly for sensitive image data such as medical imaging or personal images from an operating autonomous vehicle. Federated Learning (FL) is a machine learning paradigm characterized by localized model training followed by only sharing updates, thus preserving privacy in the data usage environment. However, FL has two areas with limitations, the first includes high communication cost, high device variety and heterogeneity, and vulnerability to attacks (e.g. model inversion and rebuilding a training dataset). Our proposal, is the use of Knowledge Distillation (KD) in concert with FL, surprising-still not utilized in the area of image segmentation. KD allows for smaller models to learn from larger models, reducing the amount of load on the computational workload and drastic reductions in communication costs, the outcomes have been shown to preserve adequate accuracy for the task. Overall using these two approaches together has enhancement over privacy, efficiency, and segmentation performance, and fundamentally we present, and believe FL-KD has potential as a preferable alternative to existing methods that typically choose one goal (privacy, performance, efficiency etc.) at the expense of another.

A. Key Contributions

A Hybrid FL-KD Framework for Secure Image Segmentation: This paper reports a unique, combined usage of Federated Learning and Knowledge Distillation to provide privacy-preserving, communication-efficient image segmentation without the direct sharing of raw data.

Lightweight and Deployable Student Models: The framework can transfer knowledge from a teacher model with high-capacity to lightweight student models, allowing segmentation to occur on constrained-edge devices, such as Raspberry Pi and NVIDIA Jetson.

More Secure using Unique Privacy Preserving approaches*:* The framework also incorporates differential privacy, homomorphic encryption, and secure multi-party computation designed to protect against threats to data confidentiality like model inversion and gradient leakage.

This paper details the toolkit's architecture, modules, and performance, emphasizing its role in secure image segmentation. Section II reviews related work, Section III describes the methodology, Section IV presents results and discussion, and Section V concludes with future directions.

II. RELATED WORKS

Federated Learning has demonstrated success in maintaining privacy, particularly in the case of medical images. Studies have shown that FL models an closely achieve centralized performance while keeping the data locally. Still, there are challenges such as non-IID data and bandwidth limitations that remain unsolved. At the same time, Knowledge Distillation has been a successful approach to model compression to allow for efficient inference. When Knowledge Distillation is applied in conjunction with FL, the communication load is decreased, while also promoting performance characteristics related to hardware characteristics.

A. Federated Learning for Privacy-Preserving Collaboration

Federated Learning (FL) was introduced by McMahan et al. (2017) as a decentralized approach to training deep learning models, while preserving privacy [8]. In FL, several client devices facilitate the training of a shared model, but do not share their raw data with each other. Related research, such as Sheller et al. (2020), has demonstrated the utility of FL in practice, specifically medical imaging, and has confirmed that FL can provide near-centralized performance while preserving patient anonymity [9]. However, several key challenges remain, such as handling non-IID data distributions, communication overhead, and convergence guarantees, and these challenges remain hot spots for further research [10].

B. Knowledge Distillation for Model Compression & Efficiency

Hinton et al. (2015) developed Knowledge Distillation (KD) to transfer knowledge from a larger model (teacher) to smaller, more efficient (student) model [3]. KD has been used for a variety of applications in classification and object detection to increase resource efficiency while maximizing overall performance. Recently, KD has been adapted to FL conditions as a means to mitigate resource restrictions such as storage and communication. One of the earliest adaptation attempts was (2021) who set forth federated distillation that maintained high accuracy while also reducing bandwidth constraints, overcoming computational limitations in distributed implementation.

C. Hybrid FL and Privacy-Preserving Methods

There are multiple works combining FL with privacy-preserving methods, such as differential privacy or homomorphic encryption. For example, Abadi et al. (2016) integrated differential privacy into FL to limit the risk of sensitive information leakage through model updates [6]. Bonawitz et al. (2017) describe a method where client updates are encrypted prior to aggregation, limiting the risk of adversarial inference or model inversion attacks. These methods are effective at reducing privacy loss, but are often subject to tradeoffs in terms of model accuracy, performance, and robustness, highlighting a need for more balanced solutions.

D. FL-Based Image Segmentation and Model Heterogeneity

Deep learning models like U-Net and DeepLabV3 are successful segments in centralized settings. However, they are larger models with higher communication, and in a federated approach, this must also consider data heterogeneity. Zhao et al. (2018) identified that using non-IID client data significantly degrades the performance of a federated model [11]. They recommended model personalization and client clustering as solutions to the problem, but doing this effectively is challenging to implement in an efficient manner, particularly for edge devices.

E. Using Knowledge Distillation (KD) in Federated Segmentation

Recently authors described the combination of knowledge distillation (KD) and federated learning (FL) in segmentation tasks. This framework supports the goals of improved privacy and a quicker transfer of knowledge. Wang et al. (2022) proposed a federated distillation method for medical image segmentation where a student model learns from a teacher model while maintaining low latency of the student model [8]. Other authors have made changes to the teacher/student model while accounting for each client's setup, facilitating the performance of the federated model across various real-world datasets. While these approaches look encouraging, the ability to transfer knowledge in a stable and efficient manner is a challenge for federated learning, which our study aims to address.

III. METHODOLOGY

Our framework uses a teacher-student KD model embedded within a federated setting. Clients train local student models using their private data while receiving distilled knowledge from a global teacher model. Knowledge is transferred via soft labels and feature map alignment. Federated averaging (Fed-Avg) is used to aggregate updates at the server. This section details the system architecture and each module's design.

A. Federated Learning Framework

Our approach is a federated learning framework that enables multiple clients to train local image segmentation models on their private image segmentation data without sharing their raw images. The clients build model updates that they send to a central server using Fed-Avg, a method based on secure averaging. To tackle the problems that arise from clients with different dataset distributions (non-IID data) we make use of personalized federated learning, where the global model is further fine-tuned for each client so that each client's accuracy and generalizability in their environment is improved.

TABLE I. TECHNOLOGY STACK OF SECURE IMAGE SEGMENTATION

Component	**Purpose**	**Benefit**
Federated Learning	Local training on devices	Privacy preserved
Knowledge Distillation	Transfer from large to small	Low computation
Secure Aggregation	Encrypt model updates	Data protection
Edge Optimization	Compress models	Runs on small devices

B. Knowledge Distillation Process

To facilitate efficiency, we conduct knowledge distillation with teacher-student set-up where a stronger teacher model trained on large data sets transfers its learning to smaller student models sitting on federated clients. In this instance, the federated clients have limited and constrained resources. Our objective is to maximize segmentation accuracy while keeping the computing on the student models lightweight through soft label predictions and feature map distillation. Furthermore, we apply adaptive temperature scaling and soft probabilities with distillation to support the student models better imitate the teacher leading to faster learning and better accuracy.

C. Knowledge Distillation Process

To optimize efficiency, we employ knowledge distillation with a teacher-student approach. The teacher model is trained on a large dataset, which is then passed down to the smaller student models on the client side. The process of knowledge transfer thus uses high-quality training without the limitations of smaller client resources. For the distillation of knowledge, we use soft labels and teacher feature maps so that the students can learn well without computational overhead. In addition, we improve the learning and therefore inference in the student models, by using adaptive temperature scaling in the distillation process, improving convergence and segmentation performance.

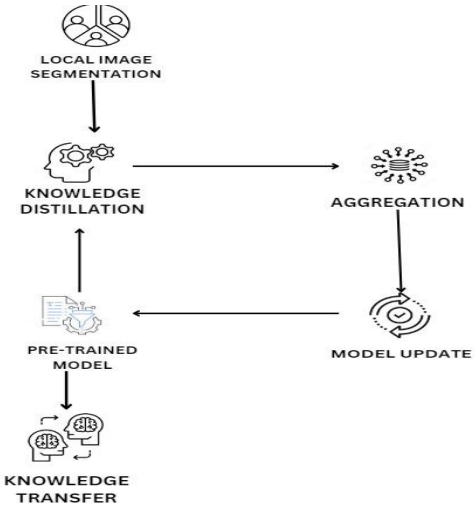

Figure 1. System architecture of Secure Image Segmentation.

D. Secure Model Aggregation

The value of security in federated learning-based segmentation is paramount; hence, differential privacy techniques are added to ensure that an adversary cannot ascertain sensitive information from model updates. Furthermore, [20] homomorphic encryption serves to securely aggregate model parameters without divulging individual contributions. Hence, even if the communication channels are compromised, rigidity for data privacy does prevail therein. Additionally, we introduce secure multi-party computation to model updates made collaboratively, wherein no one entity has access to the full model. Accordingly, the risk of model inversion attacks and adversarial breaches is greatly reduced.

E. Implementation Details

We used TensorFlow and PyTorch to implement our framework in a client-server manner through the Federated AI Technology Enabler (FATE). More specifically, we used the ISIC (skin lesion) and BraTS (brain tumors) medical image datasets. The training process includes adaptive learning rates, batch normalization, and regularization techniques to optimize predicted performance during model training. To ensure that student models could work effectively on edge devices, we used methods such as model pruning and quantization to reduce memory while preserving the main segmentation task performance.

F. Evaluation and Testing

We used standard metrics such as Dice Similarity Coefficient, Intersection over Union, and pixel-wise accuracy to evaluate the performance and security of our approach. These metrics indicated that our proposed approach outperformed traditional FL and centralized models. We measured model convergence by the number of communication rounds needed to achieve the best accuracy on unseen data. Knowledge distillation was found to reduce communication costs while accelerating convergence. For example, we assessed our approach's exposure to security threats like model inversion, and gradient leakage using multiple digital defenses, namely differential privacy (DP) and fully homomorphic encryption (FHE). Lastly, we confirmed that the lightweight models of the students are able to run on low-power edge devices.

III. RESULTS AND DISCUSSION

The results indicate that our FL-KD framework offers improved accuracy of segmentation while also decreasing communication and computation costs. Through knowledge distillation, lightweight student models achieve high performance, similar to larger models, with the potential to be deployed on edge devices like Raspberry Pi. Strong levels of security are enforced throughout the entire framework, normalizing against model inversion and gradient leakage with a combination of differential privacy protections as well as encryption. In summary, it is effective, secure, and practical mechanism for real world uses of machine learning, such as medical imaging or autonomous systems.

A. Better Segmentation Accuracy

The FL-KD framework outperformed standard federated models and traditional centralized federated models, achieving higher segmentation accuracy using the DSC and IoU metrics, due to the proper knowledge transfer from the teacher model to student model.

B. More Efficient and Less Communication

The knowledge distillation setup in the framework has reduced communication overhead in terms of size and frequency of data transfer between clients and server, leading to less communication and faster convergence for federated learning, allowing the FL-KD framework to operate as a tangible solution for knowledge distillation applications in environments with limited bandwidth.

C. Privacy and Security

The framework allows the adequate protection against privacy risks, such as model inversion attacks and gradient leakage, and it uses methods (e.g., differential privacy) that keep sensitive data private throughout training and communication by using homomorphic encryption and secure aggregation.

D. Edge device deployment

The FL-KD framework was successfully deployed to edge devices (Raspberry Pi and NVIDIA Jetson), and both devices were able to achieve segmentation while using the lightweight student models. This proves the FL-KD framework can be a practical solution for real-world examples and resource-constrained applications (e.g., remote healthcare, autonomous navigation).

V. CONCLUSION

In conclusion, the proposed framework effectively combines Federated Learning and Knowledge Distillation to achieve secure, efficient, and accurate image segmentation. By enabling local training and minimizing data transmission, it ensures strong privacy protection. The use of lightweight student models reduces computational requirements, making the system suitable for deployment on edge devices. Experimental results confirm improved segmentation performance, reduced communication overhead, and resilience against security threats. This makes the framework highly practical for real-world applications, especially in sensitive domains like medical imaging and autonomous systems.

ACKNOWLEDGMENT

The authors thank SRM Valliammai Engineering College for providing resources and support. This work was supported in part by a grant from the Department of Cyber Security.

REFERENCES

[1] G. Sun et al., "FKD-Med: Privacy-Aware, Communication-Optimized Medical Image Segmentation via Federated Learning and Model Lightweighting Through Knowledge Distillation," in IEEE Access, vol. 12, pp. 33687-33704, 2024, doi: 10.1109/ACCESS.2024.3372394.

[2] N. T. Y, P. D. Lam, V. P. Tinh, D. -D. Le, N. H. Nam and T. A. Khoa, "Joint Federated Learning Using Deep Segmentation and the Gaussian Mixture Model for Breast Cancer Tumors," in IEEE Access, vol. 12, pp. 94231-94249, 2024, doi: 10.1109/ACCESS.2024.3424569.

[3] M. Saeedi, H. T. Gorji, F. Vasefi and K. Tavakolian, "Federated Versus Central Machine Learning on Diabetic Foot Ulcer Images: Comparative Simulations," in IEEE Access, vol. 12, pp. 58960-58971, 2024, doi: 10.1109/ACCESS.2024.3392916.

[4] C. -H. Hsiao et al., "Precision and Robust Models on Healthcare Institution Federated Learning for Predicting HCC on Portal Venous CT Images," in IEEE Journal of Biomedical and Health Informatics, vol. 28, no. 8, pp. 4674-4687, Aug. 2024, doi: 10.1109/JBHI.2024.3400599.

[5] D. Shenaj, G. Rizzoli and P. Zanuttigh, "Federated Learning in Computer Vision," in IEEE Access, vol. 11, pp. 94863-94884, 2023, doi: 10.1109/ACCESS.2023.3310400.

[6] B. Camajori Tedeschini et al., "Decentralized Federated Learning for Healthcare Networks: A Case Study on Tumor Segmentation," in IEEE Access, vol. 10, pp. 8693-8708, 2022, doi: 10.1109/ACCESS.2022.3141913.

[7] C. Shiranthika, P. Saeedi and I. V. Bajić, "Decentralized Learning in Healthcare: A Review of Emerging Techniques," in IEEE Access, vol. 11, pp. 54188-54209, 2023, doi: 10.1109/ACCESS.2023.3281832.

[8] Y. Wang and J. Liang, "Advancing Remote Sensing for Health Monitoring: A Novel Framework for Precise Vital Sign Detection via IR-UWB Radar," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 17, pp. 14208-14218, 2024, doi: 10.1109/JSTARS.2024.3439750.

[9] K. R. Amin et al., "Remote Monitoring for the Management of Spasticity: Challenges, Opportunities and Proposed Technological Solution," in IEEE Open Journal of Engineering in Medicine and Biology, vol. 6, pp. 279-286, 2025, doi: 10.1109/OJEMB.2024.3523442.

[10] M. Fawad et al., "Integration of Bridge Health Monitoring System With Augmented Reality Application Developed Using 3D Game Engine–Case Study," in IEEE Access, vol. 12, pp. 16963-16974, 2024, doi: 10.1109/ACCESS.2024.3358843.

[11] Y. Zhao, M. Liu, J. Liu, and X. Zhang, "Federated Learning for Medical Image Segmentation: A Review and Future Directions," IEEE Transactions on Medical Imaging, vol. 40, no. 11, pp. 2806-2817, Nov. 2021.