Toward Robust Privacy in IoMT and Medical Federated Learning: A Critical Review of Differential Privacy

Emran AlTamimi
Department of Computer Science
Qatar University
Doha, Qatar
ea1510662@qu.edu.qa

***Abstract*—This paper critically examines the limitations of existing privacy-preserving techniques in the domain of Internet of Medical Things utilising Federated Learning. The study highlights the insufficient evaluation of privacy guarantees, the lack of intuitive explanations for practitioners, and the absence of evaluation in non-iid data scenarios as significant challenges that need to be addressed. To overcome these limitations, a novel research idea is proposed to enhance the utility and privacy of data in FL. The intended approach involves training local models on original data, sharing only labels of requested samples, and utilizing the differentially private context space. By doing so, the aim is to improve the accuracy and relevance of predictions made by the central model while ensuring user privacy. The advantages of this framework lie in its potential for enhanced utility, reduced information exposure, and flexibility across various FL settings.**

***Index Terms*—Federated Learning, Differential Privacy, Internet of Medical Things**

I. Introduction and motivation

The growing adoption of the Internet of Medical Things (IoMT) and the increased availability of medical datasets have led to significant advancements in healthcare technology. These advancements have improved patient care, enabled remote monitoring, and facilitated personalized treatment plans [1], [2]. However, the increasing volume and sensitivity of medical data pose significant privacy and security challenges. To address these concerns, researchers and practitioners are exploring the potential of federated learning (FL) and differential privacy (DP) in developing secure and privacy-preserving medical applications.

FL enables training machine learning models on distributed datasets without exchanging data samples between clients. This approach not only preserves data privacy but also enables the development of robust models that can benefit from diverse data sources. FL research is primarily concerned with three issues: increasing efficiency and effectiveness, improving security against attacks aimed at compromising the integrity of FL models, and protecting privacy by preventing data leakage. FL systems typically have two roles: clients that hold local datasets and a server that orchestrates the entire training process and updates the global model without getting client datasets [3].

Data privacy can be further enhanced by combining FL with another privacy model called DP. DP ensures that the existence or omission of any one record in a database is not discernible from query responses, up to an exponential factor in the privacy budget parameter, ϵ. The lower the ϵ, the greater the protection. The quantity of noise necessary to meet DP is inversely correlated to ϵ and directly correlated to the query function's global sensitivity. Because of its strong privacy guarantee and composability, DP is regarded by many as the gold standard in privacy protection [4].

Integrating DP into FL can help address some of the inherent privacy and security concerns associated with medical datasets and applications.DP may prevent private patient data from being inferred by an adversary even if they possess the model updates by incorporating noise to the model updates. This combination of FL and DP can lead to robust privacy-preserving solutions that maintain high utility and accuracy while ensuring that sensitive patient data remains protected [5]–[7].

In summary, integrating IoMT, FL, and DP presents a promising direction for secure and privacy-preserving healthcare solutions. These privacy-preserving techniques are crucial for maintaining patient trust and improving healthcare services. By effectively applying these methods, healthcare systems can revolutionize the industry while mitigating risks associated with data leakage and security breaches.

The motivation behind this review lies in the growing need to protect sensitive medical data when applying FL, where information security is of utmost importance. DP offers a promising solution with strong guarantees for privacy preservation; however, there is a risk of misuse by researchers and practitioners, as highlighted by Blanco et al. [8]. Therefore, it is crucial to gain a proper understanding of how to incorporate DP in FL, particularly within the healthcare domain.

The main goal of this study is to critically assess recent and state-of-the-art approaches for enhancing privacy in FL through the use of DP in healthcare applications. While our

978-1-6654-9071-9/23/$31.00 ©2023 IEEE

work shares similarities with [8], who focused on centralized machine learning and provided a generic critique of applying DP in FL, our paper specifically targets the use of DP in FL in IoMT. Our contribution aims to offer a better understanding of applying DP in FL for healthcare machine learning researchers and practitioners, and to address the question of whether DP is a suitable approach for enhancing privacy in FL within this context.

II. The problem

In FL systems, privacy risks and attacks may occur at various components and stages, such as data poisoning, model poisoning, and data privacy leakage. Data needs to be transmitted between participants and aggregators, including the local weights or the gradients, the aggregated weights or the aggregated gradients, and the final trained model, each can be exploited for privacy leakage [9], [10]. There are multiple types of attacks that can be launched against FL systems, aiming to learn confidential information about the datasets used during training. Gradient updates and weight updates can be exploited to reveal sensitive information about the training datasets, for example, training samples and labels, membership, and class representatives [11], [12]. The centralized trained model in a FL system also poses risks, as adversaries can exploit the final model parameters or predictions to launch attacks [13]. Inference of class representatives, memberships, characteristics of training data, training samples, and labels are some of the categories under which inference attacks fall. These attacks impact the privacy of training datasets by revealing sensitive information or reconstructing examples used in the training process and their labels [14]. Potential adversaries in FL systems include participants or eavesdroppers who can launch attacks or cause privacy leakages. Privacy risks and attacks are evaluated using various metrics or methods, such as the success or effectiveness of these attacks in revealing sensitive information or reconstructing training data samples and labels [9], [15]. To prevent privacy leakages and protect FL systems from various attacks, several defense strategies and techniques are known, including perturbation-based schemes to defend against gradient leakages and other techniques to protect against different types of attacks. By understanding the risks, adversaries, and attacks in FL systems, researchers and practitioners can develop more robust defenses and enhance the security and privacy of these systems [16]. DP has gained popularity among researchers in the field due to its strong privacy guarantees, which are mostly independent of the background information of the attacker, and its convenient composability properties. DP offers a rigorous mathematical framework for protecting individual privacy in the context of data analysis and has been considered the gold standard in privacy protection. Its applicability in various domains, including machine learning, has attracted significant interest from both academia and industry, leading to numerous DP-based solutions and methodologies. However, the popularity of DP also raises concerns about its potential misuse or misinterpretation, especially when applied outside its intended scope or with unreasonable privacy parameters [8].

DP was originally intended to analyze the data without revealing sensitive information through interactive queries to a database, providing strong guarantees independent of background information any attacker might have. It necessitates the addition of noise to the query function to mask a single record's presence or absence. The privacy budget, referred to as ϵ, defines the degree of protection; lower values of ϵ offer a greater privacy guarantee. This quantity of noise included in the query function is directly proportional to the global sensitivity of the query function and inversely proportional to ϵ. DP is utilized during model training in machine learning to protect the privacy of users' data. Sequential composition is used because successive model training epochs are computed on the same or overlapping data, and this requires the total privacy budget (ϵ to be divided up for different queries. To make DP compatible with machine learning, relaxations of the DP definition are used, such as (ϵ, δ)-DP, concentrated DP, zero-concentrated DP, and Rényi DP. These loosenings aim to decrease the necessary privacy budget (ϵ and necessitate less noise addition, which permits more utility preservation at the expense of a higher probability of privacy leakage. ((ϵ, δ)-DP is the most typical relaxation, integrating the failure probability δ, which reflects the probability that strict (ϵ-DP is not satisfied. Additional relaxations, such as concentrated DP, zero-concentrated DP, and Rényi DP, emphasize on the average rather than the worst-case privacy loss. The moments accountant method was introduced to bound the cumulative privacy loss across successive epochs during machine learning training, allowing single-digit (ϵ values to be reached while still preserving meaningful utility. There are several limitations and concerns that motivate us to investigate the use of DP in FL. Throughout the literature review of the state of the art, we will delve into these concerns and examine how researchers have attempted to address them. Firstly, weak privacy guarantees often arise due to the privacy parameter (ϵ) is set too high to be considered private in FL applications. This issue warrants further exploration to ensure that strong privacy protection is provided for the clients. Secondly, the significant impact on model accuracy is a concern that merits investigation. For a given privacy level, the clients in the FL setup has a major influence on the performance, which is particularly problematic when they're below 1,000 for instance. In such cases, the impact on accuracy can be substantial, sometimes to the point where the model hardly converges [8]. Furthermore, the competitive accuracy provided by DP-FL under the assumptions of an exceedingly high number of clients and identically distributed data among them is often unrealistic. The generalizability of DP-FL is called into question as a result. Additionally, the issue of non-i.i.d. data among clients and the distinction between instance-level and client-level privacy protection add complexity to the privacy preservation process in FL. One crucial concern is the trust placed in the model manager when implementing central differential privacy (CDP). Clients are required to trust the model manager with their exact model updates, which contradicts the fundamental purpose of privacy protection. This issue necessitates the exploration of alternative approaches that mitigate the reliance on a trusted model manager. Lastly, and most importantly for this review, the limited applicability of DP-FL in smaller configurations, such as medical applications, poses a significant challenge. In these situations, addressing privacy concerns by using DP-FL can result in unacceptable accuracy loss, which has a direct impact on the quality of the developed models and their practical utility. This review will specifically focus on the literature related to medical applications, emphasizing the critical importance of finding viable solutions for preserving privacy while maintaining acceptable levels of model accuracy in these sensitive domains.

A. Literature Review

This literature review provides a comprehensive overview of the state-of-the-art research in this field, summarizing key findings from several seminal studies. The initial study in our review is by Ho et al. [17]. Their work lies at the intersection of FL, DP, and COVID-19 detection. They developed a FL model for COVID-19 detection using patient symptom information and chest X-ray images, with the model trained across multiple data sources or 'clients' (hospitals) while ensuring no data sharing among these entities. The authors incorporated DP stochastic gradient descent (DP-SGD) into their model to enhance its resilience against adaptive attacks with auxiliary information, using local DP. The DP-SGD algorithm involves several steps, including computing gradients only for a percentage of the examples, then fixing a maximum L2 norm, adding Gaussian noise to the gradients when determining the average step, and updating the model parameters by applying the learning rate to the noised and clipped gradients. The researchers paid particular attention to the issue of non-identically distributed (Non-IID) data. They found that the accuracy of the model decreased when dealing with Non-IID data, with reductions ranging from 1.28% to 55.32%, depending on the dataset and the degree of non-IIDness. However, the effect of Non-IID data on DP was not addressed. The privacy utility tradeoff was studied by increasing the privacy budget and studying the effect on privacy. On 'low' epsilons (250 and 42) the model did not even converge, the privacy budget for the optimal model was absurdly high (390 000 and 1600) privacy budget. They then experimented with only participating a portion of the consumers each round, however, they increased the number of participants. Building upon the privacy-preserving FL model proposed by Ho et al. [17], another intriguing approach was presented by Li et al. [18] focusing on the application of FL for medical image analysis, particularly brain tumor segmentation. The authors implemented a client-server architecture, maintaining a global Deep Neural Network (DNN) model on a centralized server, while clients carried out local Stochastic Gradient Descent (SGD) updates. This setup ensured the privacy of individual data as the model training occurred locally on each client, and only the model updates were shared with the server. A crucial aspect of this approach was the application of selective parameter sharing to prevent overfitting and potential memorization of local training examples, thereby reducing the risk of revealing sensitive training data. The selective parameter-sharing mechanism only allowed a fraction of the local model updates to be shared if their absolute values were greater than a certain threshold. In addition, the updates were clipped within a fixed range to ensure data privacy. To further enhance the privacy of the model, the authors incorporated the Sparse Vector Technique (SVT) into the selective parameter-sharing process. This involved introducing Laplacian noise into the selection and sharing of model updates, providing stronger DP guarantees. The noise-infused selection procedure was repeated until a predefined fraction of local model updates was released, ensuring that the entire process complied with the principles of DP. The study conducted various experiments to assess the performance of their privacy-preserving FL system. These tests explored the impact of selective parameter updates, with an emphasis on the percentage of the model shared and gradient clipping value. The researchers also studied partial model sharing, observing the system's performance with different levels of shared information. In the DP module experiment, the trade-off between privacy protection and performance was assessed by adjusting the noise level and shared parameters, the utility of the data was relatively good even with a privacy budget smaller than 1. Finally, the effectiveness of momentum restarting and weighted averaging in the FL procedure was evaluated, particularly their impact on the convergence speed and the quality of the global model.

Notably, the authors did not address the problem of Non-IID data in this article and its effects on DP. However, it's expected that partial model sharing would solve the issue.

The proposed research by the authors in [19] introduces a novel approach to augment FL with Bayesian differential privacy (BDP), a natural relaxation of DP that offers tighter guarantees. The key idea behind BDP is to leverage the knowledge that machine learning tasks are often constrained to specific types of data, and this information, along with prior data distribution, may be available to potential attackers.

Unlike traditional DP, which treats all data as equally likely and introduces substantial amounts of noise to hide differences, BDP calibrates the noise based on the underlying data distribution. As a result, BDP provides more precise privacy guarantees than traditional DP for two datasets gathered from an identical distribution and utilizing the same privacy mechanism with the same amount of noise. This approach offers potential benefits in terms of improved utility and reduced noise compared to traditional DP mechanisms. The proposed approach introduces joint accounting to address the challenge of slow convergence or divergence in FL due to added noise for privacy. Joint accounting leverages instance-level privacy by re-counting the noise already present in client updates towards client-level privacy guarantees. This eliminates the need for additional noise and allows the server to compute metrics using the existing noise. The procedure involves clients computing private outcome distributions, while the server performs averaging to calculate necessary metrics. By incorporating joint accounting, the approach achieves tight instance and client privacy guarantees while maintaining a comparable convergence speed to client privacy-only solutions. The evaluation conducted in this study demonstrates the benefits of BDP in FL, including the need for less noise to achieve privacy guarantees, tighter privacy bounds, and applicability in non-iid settings. The results show that BDP allows for faster convergence and fewer communication rounds while maintaining meaningful privacy guarantees, even with a large number of clients. By employing joint accounting, client privacy can be achieved alongside instance privacy without additional costs or compromising accuracy.

A more recent study conducted by [20] in 2023, suggested the utilization of differential privacy on the gradients of local models. Moreover, they enforce user anonymity in their architecture using single-key encryption on the client-side and double-key encryption at the edge server. An interesting feature of their proposed solution was to implement the randomized feature selection and client selection in alternate iterations. It was found that decreasing the number of clients or features might increase the system performance in terms of accuracy, precision, F1 score, and Matthew's correlation coefficient (MCC), but it also increases the chances of model overfitting. Hence, they suggested selecting at least 20% of clients and 50% of attributes to avoid overfitting. Then they tested the performance by applying differential privacy to the gradients of model. By inducing high noises where λ=0.08 the model performed well with the accuracy of 94%. The other evaluation measures reported similar outcomes thus proving the robustness of their model Fed-Select. In another paper [21], this research was conducted over the time-series dataset for the heartbeat of patients. The researchers adopted a creative utilization of Variational Auto Encoder (VAE). They converted the medical dataset into vectors and then passed it through VAE. The degenerated dataset, (with comparatively lower quality than the original dataset) was then used to train a long short-term memory network, LSTM model. They reported that the probability of data reconstruction drops by 25% as compared to [22]. But in their paper, they didn't investigate the impact of loss function on the data reconstruction. We assume that, with each epoch the training loss will decrease hence at certain point the data reconstruction will be easier.

III. Potential Research Project

A. Problem statement of the research project idea(s)

- One of the main limitations we found in the literature is the lack of regirous evaluatoin for the gaurentees of privacy offered by DP. This issue was also highlighted in the context of centralized machine learning in [].
- Another issue is the lack of intuitive explanation offered in the literature for practitioners. Since applying DP in different ways means different things, a lack of intuitive explanation of what the privacy budgets mean in each specific framework might lead to mis-use and leakage of information. In other words, the original DP budget, as explained in [] quantifies the upper bound on the probability of a significant change in the output due to the inclusion or exclusion of any individual data point. However, when DP is applied to the weights of the model, epsilon does not directly capture how much the query or weights are off but rather reflects the overall privacy level and the impact on the learning process. That is done by limiting the overall effect of a single data point on the model's weights or gradients.
- There's no evaluation of the effect of DP when applied to non-IID data. Take the extreme case of two users, one with diabetes and one without. If both users train a model with DP, their models will be differentially private with their own distribution of the data. In this case, for the diabetic user, for example, the amount of noise added will be minimal since each data point is not sensitive with respect to his own data. However, in the larger context of multiple datasets and users, his dataset is very sensitive, which was not captured in his calculation of the sensitivity of his data. Therefore, the central model can infer that a user is diabetic even when applying DP.

B. Description of the research idea

In this section, we present a novel research idea aimed at enhancing the utility and privacy of data in FL scenarios. Our approach involves training local models on original data and sharing only the labels of requested samples with the central model, along with the corresponding DP context space. This context space represents the distribution of the data each local model was trained on, allowing the central model to identify the most relevant models for prediction based on similar data distributions. Our research idea focuses on improving the utility and privacy aspects of FL by following a step-by-step process. The main steps involved in our approach are as follows:

1) Training Each Model on Original Data: Each user participating in the FL process trains a local model using their original data without applying any privacy-preserving mechanisms. This ensures that the models capture the unique characteristics and nuances of each user's data.
2) Sharing Labels of Requested Samples: Instead of transmitting the entire model or confidence scores, only the labels of the requested samples are shared between the local and central models. By sharing labels alone, we minimize the risk of exposing sensitive information or revealing specific model parameters.
3) Sharing DP Context Space: In addition to the labels, each local model also shares its DP context space with the central model. The context space represents the distribution of the data that the local model was trained on. For instance, it includes the ranges or values of different features. This information provides insights into the data distributions encountered during local model training.
4) Context Space Comparison and Model Selection: Using the shared context spaces, the central model performs a comparison to identify the local models that have

similar data distributions. By determining the closest context spaces to the samples of interest, the central model selects the most relevant local models for further prediction.

5) Requesting Predictions from Selected Models: Once the local models with similar data distributions are identified, the central model requests predictions for the specific samples of interest from these selected models. By leveraging models trained on similar data distributions, we aim to enhance the accuracy and utility of the predictions made by the central model.

The proposed framework has several advantages over the literature which are highlighted below:

- Enhanced Utility: By selecting local models trained on similar data distributions, our approach aims to improve the accuracy and relevance of predictions made by the central model. This can potentially lead to better overall utility and performance in FL scenarios.
- Reduced Information Exposure: By sharing only labels and context spaces, our approach mitigates the risk of revealing sensitive information or specific model parameters during the FL process. This helps maintain user privacy and confidentiality.
- Flexibility and Compatibility: Our research idea can be applied in various FL settings, accommodating different data distributions and characteristics. It can also be integrated with existing privacy-preserving mechanisms, such as DP, to further enhance privacy guarantees.

However, there are still some limitations that must be addressed such as:

- Context Space Accuracy: The effectiveness of our approach heavily relies on the accurate representation of the context space shared by each local model. Inaccurate or imprecise context spaces may lead to suboptimal model selection and potentially compromise the utility of the central model's predictions.
- Privacy-Utility Trade-off: While our approach aims to balance utility and privacy, there may still be a trade-off between these two factors. Sharing context spaces could potentially reveal some information about the data distributions, posing a privacy risk. Striking an optimal balance between utility and privacy remains a challenge.

C. Evaluation plan

To evaluate the proposed solution and assess its effectiveness, we outline a comprehensive evaluation plan that encompasses experimental design, datasets, and performance metrics. Our evaluation plan particularly focuses on assessing the impact of the proposed model on privacy and performance trade-offs. The following plan provides a framework for evaluating the effectiveness of our research idea:

- Dataset Selection: We opt for the APTOS 2019 Blindness Detection Challenge dataset, because it has 4 classification labels and a low number of samples (3662 images). This dataset reflects a realistic scenario of several clinicians cooperating to build a more powerful model.
- Privacy Mechanism Selection: Incorporate DP with the Laplacian mechanism and (ϵ- differential privacy), into the proposed model when sharing the context spaces.
- Baseline Comparison: Establish appropriate baseline models for comparison. Consider traditional FL approaches without privacy guarantees, as well as other relevant privacy-preserving FL techniques.

The evaluation aims to demonstrate the effectiveness of the proposed solution in achieving a balance between privacy and performance in FL. By evaluating the proposed model based on the outlined experimental plan, we can assess its impact on privacy and performance trade-offs in the following ways:

- Privacy Assessment: The evaluation metrics, such as privacy loss or inference attacks, give a quantitative indication of the level of privacy that the suggested model has achieved. Comparisons with baselines and different privacy budget settings allow us to understand the level of privacy preservation and the trade-offs involved. In particular, we aim to train an adversarial model to understand our model's effectiveness against the membership inference attack and the model inversion attack.
- Utility Evaluation: The utility metrics assess the performance of the central model's predictions, considering accuracy, precision, recall, F1 score, or AUC. Comparing the utility achieved by the proposed model with the baselines helps determine the impact of privacy-preserving mechanisms on the model's effectiveness.
- Trade-off Analysis: By analyzing the results obtained from privacy and utility evaluations, we can assess the trade-off between privacy and performance. This analysis provides insights into the practicality and effectiveness of the proposed solution in achieving a desirable balance between preserving privacy and maintaining useful predictions.

Overall, the evaluation demonstrates the impact of the proposed solution on privacy and performance trade-offs. It validates the effectiveness of the proposed model in maintaining a reasonable level of privacy while achieving utility comparable to or better than baseline approaches. The results obtained through the evaluation provide evidence of the feasibility and potential benefits of the proposed solution in real-world FL.

IV. Conclusion

In conclusion, our review highlights several limitations in the literature regarding privacy-preserving techniques, specifically in the context of DPO and FL. Insufficient evaluation of privacy guarantees, lack of intuitive explanations for practitioners, and the absence of evaluation in non-iid data scenarios are notable challenges that must be addressed.

To address these limitations, we propose a novel research idea that focuses on enhancing the utility and privacy of data in FL. By training local models on original data, sharing only labels of requested samples, and utilizing the DPO context space, our approach aims to improve the accuracy and relevance of predictions made by the central model while preserving user privacy.

The advantages of our proposed framework lie in its potential for enhanced utility, reduced information exposure, and flexibility across various FL settings. However, it is crucial to acknowledge the limitations related to context space accuracy and the inherent trade-off between utility and privacy.

Continued research and development in privacy-preserving techniques for IoMT and medical applications are of paramount significance. Further investigation is needed to address the limitations identified in the literature and evaluate the effectiveness and practicality of the proposed approaches. The development of more rigorous evaluation methodologies, intuitive explanations for practitioners, and considerations for non-iid data scenarios will contribute to the advancement and wider adoption of privacy-preserving techniques in these critical domains.

By addressing these challenges and pursuing further research, we can ensure the effective protection of sensitive medical data while enabling the utilization of advanced machine learning techniques for improved healthcare outcomes.

References

[1] Y. Zhang, D. He, M. S. Obaidat, P. Vijayakumar, and K.-F. Hsiao, "Efficient identity-based distributed decryption scheme for electronic personal health record sharing system," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 2, pp. 384–395, 2020.

[2] V. Pandi, P. Perumal, B. Balusamy, and M. Karuppiah, "A novel performance enhancing task scheduling algorithm for cloud-based e-health environment," *International Journal of E-Health and Medical Communications (IJEHMC)*, vol. 10, no. 2, pp. 102–117, 2019.

[3] X. Yin, Y. Zhu, and J. Hu, "A comprehensive survey of privacy-preserving federated learning: A taxonomy, review, and future directions," *ACM Computing Surveys (CSUR)*, vol. 54, no. 6, pp. 1–36, 2021.

[4] C. Dwork, "Differential privacy," in *Automata, Languages and Programming: 33rd International Colloquium, ICALP 2006, Venice, Italy, July 10-14, 2006, Proceedings, Part II 33*. Springer, 2006, pp. 1–12.

[5] H. B. McMahan, D. Ramage, K. Talwar, and L. Zhang, "Learning differentially private recurrent language models," *arXiv preprint arXiv:1710.06963*, 2017.

[6] V. Pihur, A. Korolova, F. Liu, S. Sankuratripati, M. Yung, D. Huang, and R. Zeng, "Differentially-private" draw and discard" machine learning," *arXiv preprint arXiv:1807.04369*, 2018.

[7] Y. Zhu, X. Yu, Y.-H. Tsai, F. Pittaluga, M. Faraki, Y.-X. Wang *et al.*, "Voting-based approaches for differentially private federated learning," *arXiv preprint arXiv:2010.04851*, 2020.

[8] A. Blanco-Justicia, D. Sánchez, J. Domingo-Ferrer, and K. Muralidhar, "A critical review on the use (and misuse) of differential privacy in machine learning," *ACM Computing Surveys*, vol. 55, no. 8, pp. 1–16, 2022.

[9] M. Fredrikson, S. Jha, and T. Ristenpart, "Model inversion attacks that exploit confidence information and basic countermeasures," in *Proceedings of the 22nd ACM SIGSAC conference on computer and communications security*, 2015, pp. 1322–1333.

[10] H. Yu, S. Yang, and S. Zhu, "Parallel restarted sgd with faster convergence and less communication: Demystifying why model averaging works for deep learning," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, 2019, pp. 5693–5700.

[11] M. Nasr, R. Shokri, and A. Houmansadr, "Comprehensive privacy analysis of deep learning: Passive and active white-box inference attacks against centralized and federated learning," in *2019 IEEE symposium on security and privacy (SP)*. IEEE, 2019, pp. 739–753.

[12] H. Li and T. Han, "An end-to-end encrypted neural network for gradient updates transmission in federated learning," *arXiv preprint arXiv:1908.08340*, 2019.

[13] D. Enthoven and Z. Al-Ars, "An overview of federated deep learning privacy attacks and defensive strategies," *Federated Learning Systems: Towards Next-Generation AI*, pp. 173–196, 2021.

[14] Z. Wang, M. Song, Z. Zhang, Y. Song, Q. Wang, and H. Qi, "Beyond inferring class representatives: User-level privacy leakage from federated learning," in *IEEE INFOCOM 2019-IEEE conference on computer communications*. IEEE, 2019, pp. 2512–2520.

[15] G. Ateniese, L. V. Mancini, A. Spognardi, A. Villani, D. Vitali, and G. Felici, "Hacking smart machines with smarter ones: How to extract meaningful data from machine learning classifiers," *International Journal of Security and Networks*, vol. 10, no. 3, pp. 137–150, 2015.

[16] D. Liu, T. Miller, R. Sayeed, and K. D. Mandl, "Fadl: Federated-autonomous deep learning for distributed electronic health record," *arXiv preprint arXiv:1811.11400*, 2018.

[17] T.-T. Ho, K.-D. Tran, and Y. Huang, "Fedsgdcovid: Federated sgd covid-19 detection under local differential privacy using chest x-ray images and symptom information," *SENSORS*, vol. 22, no. 10, MAY 2022.

[18] W. Li, F. Milletari, D. Xu, N. Rieke, J. Hancox, W. Zhu, M. Baust, Y. Cheng, S. Ourselin, M. J. Cardoso, and A. Feng, "Privacy-preserving federated brain tumour segmentation," in *MACHINE LEARNING IN MEDICAL IMAGING (MLMI 2019)*, ser. Lecture Notes in Computer Science, H. Suk, M. Liu, P. Yan, and C. Lian, Eds., vol. 11861, 2019, pp. 133–141, 10th International Workshop on Machine Learning in Medical Imaging (MLMI) / 22nd International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), Shenzhen, PEOPLES R CHINA, OCT 13-17, 2019.

[19] A. Triastcyn and B. Faltings, "Federated learning with bayesian differential privacy," in *2019 IEEE INTERNATIONAL CONFERENCE ON BIG DATA (BIG DATA)*, ser. IEEE International Conference on Big Data, C. Baru, J. Huan, L. Khan, X. Hu, R. Ak, Y. Tian, R. Barga, C. Zaniolo, K. Lee, and Y. Ye, Eds. IEEE Comp Soc; IEEE; Baidu; Very; Ankura, 2019, pp. 2587–2596, iEEE International Conference on Big Data (Big Data), Los Angeles, CA, DEC 09-12, 2019.

[20] A. K. Nair, J. Sahoo, and E. D. Raj, "Privacy preserving federated learning framework for iomt based big data analysis using edge computing," *Computer Standards & Interfaces*, p. 103720, 2023.

[21] X. Wang, J. Hu, H. Lin, W. Liu, H. Moon, and M. J. Piran, "Federated learning-empowered disease diagnosis mechanism in the internet of medical things: From the privacy-preservation perspective," *IEEE Transactions on Industrial Informatics*, 2022.

[22] R. Zhou, X. Li, B. Yong, Z. Shen, C. Wang, Q. Zhou, Y. Cao, and K.-C. Li, "Arrhythmia recognition and classification through deep learning-based approach," *International Journal of Computational Science and Engineering*, vol. 19, no. 4, pp. 506–517, 2019.