Abstract

Maintaining patient privacy in medical imaging data is critical at all stages of deep learning model development and deployment. This paper examines the tradeoff between privacy and performance in deep learning models for medical image analysis, focusing on differential privacy (DP) as a technique for safeguarding patient confidentiality. Non-differential private and differentially private models were evaluated for a chest radiography pneumonia detection task, hypothesizing that DP models could achieve similar performance to non-DP models while being less vulnerable to privacy breaches (via membership inference attacks). DP models were trained with a range of noise multiplier hyperparameter settings, which affect privacy. Testing accuracy, privacy guarantee factor (epsilon), and resulting susceptibility to membership inference attacks were evaluated. Results showed that increasing the privacy of a model can impact its performance, but optimal hyperparameter choices can potentially yield high performance while being less susceptible to attacks. The DP model with a noise multiplier of 0.02 achieved comparable testing accuracy to the non-DP model while being less prone to membership inference attacks. This project's contribution is twofold: first, training deep learning models for medical image analysis in a private setting, and second, using membership inference attack scores as a novel metric to measure the privacy of these deep learning models. This approach allows us to measure the extent to which a model's private information is disclosed to an attacker and provides a new way of evaluating the privacy of deep learning models for medical image analysis. Future research should explore various model architectures and hyperparameters that best balance performance and privacy in medical imaging based on these findings.

Introduction

Deep learning techniques are increasingly utilized in healthcare for various applications, including disease detection such as cancer, diabetes, and dementia [1], as well as medical image segmentation tasks like liver segmentation [2] and knee cartilage segmentation [3]. However, patient privacy remains a significant concern for deep learning-based algorithms, mainly due to the highly sensitive and confidential information found in medical health records [4]. These deep learning models are subject to ethical and privacy requirements set by HIPAA (Health Insurance Portability and Accountability Act) compliance guidelines, ensuring patient health information's confidentiality, integrity, and availability while protecting individuals' privacy [5]. HIPAA establishes standards and regulations for healthcare providers, health plans, and healthcare clearinghouses to safeguard protected health information (PHI). Nevertheless, these requirements may limit data sharing and hinder the development of Artificial Intelligence (AI) centric algorithms for medical tasks. Thus, it is crucial to ensure that patient privacy is carefully considered during any stage of deep learning model development and deployment.

Privacy-Preserving Machine Learning is an emerging field within Artificial Intelligence that is being increasingly studied to address this issue. Techniques such as Federated Learning, which enables the training of AI models on distributed datasets that cannot be directly accessed, Differential Privacy, which provides formal, mathematical guarantees around privacy preservation when publishing results (either directly or through AI models), and Encrypted Computation, which allows for machine learning to be performed on encrypted data, are all being applied to enhance privacy while also improving data utilization [6].

This paper aims to investigate the applicability of Differential Privacy for medical image analysis tasks while exploring its potential benefits and limitations. Differential privacy is a powerful technique that guarantees the privacy of each individual in the dataset, even if an attacker has prior knowledge of the dataset. The main concept behind differential privacy is to ensure that the results of computations on the data remain the same, whether or not the data of any one individual is included. As a result, the privacy of each individual is protected, and the output of the model cannot be used to trace back to an individual's record [7].

Furthermore, this paper evaluates the vulnerability of privately trained deep learning models to membership inference attacks in addition to exploring differential privacy algorithms. Recently developed inference attack algorithms suggest that deep networks can be queried by malicious parties to reconstruct images and text records [8]. Membership inference attacks aim to infer whether a particular data record was used during the training of a machine learning model. Even when the model is trained with strong privacy protection techniques, such attacks can be used to extract sensitive information about individuals [9].

In this study, we conduct experimental evaluations to determine the extent to which the proposed private approaches effectively mitigate the potential privacy leakage of medical records from these inference attacks. Essentially, our objective is to test the susceptibility of private models to these attacks and ascertain whether the concept of differential privacy can successfully reduce potential privacy risks while maintaining performance. This work specifically references a recent study that introduced a membership inference attack model designed to assess the privacy of deep learning models [10].

Theory

The key idea of differential privacy is that the output of a mechanism M must be almost the same whether it is run on D1 or D2, two datasets that differ by only one record, such as Patient 1 [11]. This principle is illustrated in Figure 1, where the output of a differentially private analysis will be approximately the same whether or not a particular individual's data is included. Differential privacy ensures that the two answers should be statistically indistinguishable, meaning that the output should not reveal whether or not Patient 1 was included in training or what it contained.

When designing differentially private algorithms, two key quantities must be considered: Epsilon and Accuracy. Epsilon is a metric of privacy loss that measures how much information is leaked about an individual's data when a differentially private algorithm is applied. The smaller the privacy loss value, the stronger the privacy protection provided by the algorithm. Accuracy measures how close the output of a differentially private algorithm is to the output of a non-private algorithm [12]. Achieving high accuracy in a differentially private setting may require sacrificing accuracy for privacy protection, and this will be further explored in the paper.

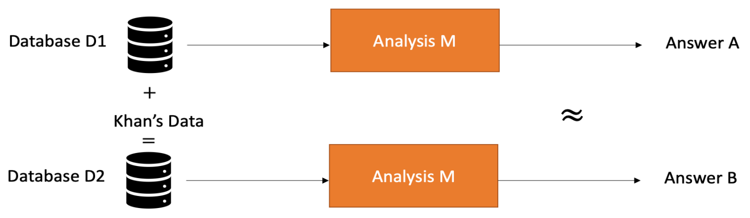

Figure 1: Comparison of computation results with and without Patient 1's data. Answer "A" is computed independently, while answer "B" is computed including Patient 1's dataset. Differential privacy guarantees indistinguishability between the two results [11]

There are several techniques to achieve differential privacy, such as Local Differential Privacy, which adds noise to the input before querying the data, and Global Differential Privacy, which adds noise to the output of the query instead of each data point [13]. Another way to achieve differential privacy is to implement DP mechanisms during model training. We utilized Differentially-Private Stochastic Gradient Descent (DP-SGD), which alters the minibatch stochastic optimization process to make it differentially private [14].

The core idea behind DP-SGD is to ensure that the model's parameter gradients preserve the differential privacy of the training data, so the resulting model will also be differentially private. This is achieved by adding noise to the parameter gradients. The Gaussian mechanism is used to determine the appropriate amount of noise to add to the training process to maintain privacy while still allowing the model to be useful. This mechanism takes into account two parameters: the noise multiplier and the bound on the gradient norm [15]. Since the gradient norm can be unbounded, especially for outliers and mislabeled inputs, the gradient norm for each sample in the batch is clipped to ensure that it does not exceed a predetermined threshold, which is referred to as the clipping threshold. This ensures that the model cannot learn more information than a set quantity from any given training sample, regardless of how different it is from the rest. This method of differentially private deep learning ensures that the privacy of each and every sample in a batch is respected by adding just enough noise to hide the largest possible gradient.

The theory underlying membership inference attacks revolves around two crucial phases: training a set of shadow models on the attacker's data and training an attack model to predict the membership of a data sample in the victim model's training dataset based on its predictions [16].

In the first phase, the attacker trains a set of shadow models using their own data to mimic the behavior of the victim model. These shadow models are then utilized to generate a dataset consisting of predictions and labels that the victim model would have produced on its own data.

In the next stage, the adversary trains an attack model to infer whether a given data sample belongs to the training dataset of the victim model by analyzing its predictions. The attack model takes the predictions made by the victim model as input and produces a probability score indicating the likelihood of the sample belonging to the victim model's training dataset. The attacker's objective is to achieve high accuracy on this dataset, referred to as the MIA (Membership Inference Attack) accuracy score, which signifies the effectiveness of the attack model in distinguishing between samples in the victim model's training dataset and those that are not [17]. It is expected that models incorporating stronger privacy protection mechanisms would be less susceptible to such attacks, resulting in lower MIA accuracy scores.

In this paper, we explore these MIA accuracy scores through experimentation to assess the impact of stronger privacy protection mechanisms on the vulnerability of models to such attacks.

Methods

The primary objective of this study was to assess the feasibility of using differentially private neural networks for medical image analysis tasks. We focused on the binary classification of X-ray images with and without pneumonia using the PneumoniaMNIST dataset. To achieve this goal, a series of experiments were conducted, including training a baseline non-differentially private model and training a range of differentially private models using varied DP-specific hyperparameter settings. The performance of these models was evaluated by comparing their accuracy to perform a tradeoff analysis between accuracy and privacy. Furthermore, membership inference attacks were conducted to assess the privacy of the differentially private models and evaluate the effectiveness of the proposed private approaches in mitigating the potential privacy leakage of medical records.

Data

In this project, we utilized the PneumoniaMNIST dataset [18], which is an open-source dataset obtained from Kaggle [19]. The dataset had already undergone pre-processing steps, including standardization, label annotation, and the removal of low-quality scans [18], so no additional pre-processing was necessary. Initially, the dataset consisted of 5,863 X-ray scan images, with 1,349 categorized as normal and 3,883 as pneumonia cases. To address the class imbalance, augmentation techniques were applied specifically to the images in the normal class. Consequently, a total of 3,059 samples depicting pneumonia and 2,989 samples representing the normal class were used for further analysis and training.

To load the data, we used Keras data loaders with a batch size of 10. The dataset had already been pre-divided into train and test folders, with approximately 5,425 images used for training and the remaining images reserved for testing.

Before training the model, we partitioned the images into separate test and train sets. For each iteration, the same data was used to ensure result reproducibility [20]. Additionally, we hot encoded the labels to represent the categorical classes: pneumonia and not-pneumonia.

Model Training

The training process for the models involved several steps. Initially, we trained a baseline non-differentially private model, followed by a series of differentially private models with varied hyperparameter settings. Both models utilized the DenseNet121 architecture, which had been pre-trained on ImageNet weights due to its excellent previous performance with medical images [21]. Training was conducted on an NVIDIA Tesla V100 GPU using the TensorFlow framework.

For the non-differentially private (non-DP) model, we employed the Adam optimizer with a learning rate of 0.0001 and a decay of 0.00001. In contrast, the differentially private (DP) models utilized the Differentially Private Stochastic Gradient Descent (DPSGD) optimizer. DPSGD, imported using the TensorFlow Privacy framework, offered two hyperparameters controlling privacy levels: the maximum gradient norm and the noise multiplier. We set the maximum gradient norm to 1.5 for optimal results, while varying the noise multiplier between 0.001 and 0.05. The training loss metric was CategoricalCrossentropy, and to prevent overfitting, an EarlyStopping callback was implemented, terminating training if the loss did not improve after three consecutive epochs. A validation split of 0.1 was employed to evaluate model performance during training.

Following the established settings, the target model was trained using the Keras fit_generator function for 25 epochs. Subsequently, we evaluated the model separately on the testing dataset to calculate its accuracy score.

To assess the robustness of the models, we conducted membership inference attacks (MIA) using the MIA library, a Python library designed for running membership inference attacks against machine learning models [22]. These attacks aim to determine whether a particular data point was used in training a model. The MIA attacks followed the implementation described in [23].

To conduct the membership inference attacks, we created two model functions: the target model and the attack model. The target model utilized the DenseNet121 architecture with the aforementioned settings, serving as the basis for training both differentially private (DP) and non-DP models. The attack model consisted of three dense layers with ReLU activations, a dropout layer, and a sigmoid activation layer. It was compiled using the Adam optimizer and binary cross-entropy loss.

After completing the training of the target model, we computed the accuracy score and epsilon value. The epsilon value was calculated using the compute_dp_sgd_privacy function from the TensorFlow Privacy library.To train the shadow models, we utilized the ShadowModelBundle from the MIA library [22]. We selected a subset of 200 samples from the test dataset for training, and the num_models parameter was set to 5.

Subsequently, we created an AttackModelBundle [22] and used the output of the shadow models to train the attack models. These attack models aimed to predict whether a given data point belonged to the dataset or not. We evaluated the attack models on the test dataset and calculated the MIA accuracy score based on the number of correct predictions made. To ensure repeatable results and account for randomness, we repeated this entire process five times for each noise multiplier, computing the average accuracy and attack accuracy.

Evaluation

We evaluated the models based on their accuracy on unseen test data and also considered their epsilon value, which represents the privacy loss metric, and membership inference attack (MIA) accuracy scores. The MIA accuracy scores assessed the effectiveness of an attack model in determining whether a specific data point belonged to the training dataset used to train the victim model. They measured the attack model's ability to distinguish between samples that were part of the victim model's training dataset and those that were not, indicating the attack model's accuracy in guessing their membership. A low membership inference attack score indicated that the model was less vulnerable to attacks attempting to reveal if a particular data point was part of the training dataset. In other words, a lower score indicated that the model better protected the privacy of individual data points and was less likely to reveal information about the training dataset. This isparticularly important in the field of medical imaging, where maintaining confidentiality and protection of patient data is crucial. By reducing the vulnerability of a model to membership inference attacks, we could better preserve the privacy of patient data.

We conducted hypothesis testing to evaluate the results, investigate the potential trade-off between accuracy and privacy, and explore the impact of privacy on membership inference attack accuracy. We tested two null hypotheses: the first stating that adding more privacy does not affect model accuracy, and the second stating that adding more privacy does not affect membership inference attack accuracy.

To analyze the relationships, we computed the Pearson correlation coefficient, which quantified the strength and direction of the relationship between variables. Additionally, we calculated the p-value using a significance level of 0.05. If the p-value was lower than this predetermined threshold, it indicated that the null hypothesis should be rejected, suggesting a significant relationship between the variables.

Results

Non-differentially private and differentially private models were evaluated based on their testing accuracy, epsilon value, and membership inference attack accuracy. The initial hypothesis aimed to train accurate and robust deep learning models for medical image analysis in a private setting using differential privacy.

The baseline non-DP model achieved an average testing accuracy of 0.89, with an average MIA accuracy score of 0.602. DP models were trained five times using noise multipliers of 0.0010, 0.0025, 0.0050, 0.0060, 0.0080, 0.0100, 0.0200, 0.0300, 0.0400, and 0.0500. Average testing accuracies were 0.8835, 0.8834, 0.8844, 0.8860, 0.8677, 0.8202, 0.8462, 0.7040, 0.5066, and 0.6022, respectively. Average MIA accuracies were 0.579, 0.5845, 0.577, 0.579, 0.585, 0.58, 0.5385, 0.4695, 0.488, and 0.466, respectively. Table 1 and Table 2 compile all the results for reference.

The box plot in Figure 2 illustrates the trade-off between privacy and accuracy. This is due to the addition of random noise to the model parameters in differential privacy, where distorting the gradients by introducing more noise during the gradient update step compromises the overall performance. Figure 3 showcases the impact of privacy on MIA accuracy, demonstrating that as the noise multiplier increases and privacy improves, the attack accuracy decreases. This validates the claim of private models' effectiveness against membership inference attacks. As expected, by subjecting our models to membership inference attacks, it was discovered that as the model becomes more private, it becomes increasingly challenging for the attack to determine if a given sample was part of the dataset. Additionally, Figure 4 indicates the amplification of variance in membership inference attacks with higher noise multipliers, suggesting model instability beyond a certain threshold. The broader spread of the box plot in Figure 2 for higher noise multipliers also reflects the relationship between noise multiplier and testing accuracy.

Figure 5 visually represents the relationship between the noise multiplier and epsilon, a privacy metric. A smaller epsilon value corresponds to stronger privacy guarantees but may result in less accurate or useful outcomes, while a larger epsilon implies weaker privacy guarantees but potentially yields more accurate or useful results.

Furthermore, two null hypotheses were tested to examine the effects of privacy on testing accuracy and membership inference attack accuracy. The first null hypothesis stated that adding more privacy does not affect model testing accuracy, while the second null hypothesis stated that adding more privacy does not affect membership inference attack accuracy. The correlation coefficient between the noise multiplier and average model testing accuracy was found to be -0.926, with a p-value of 0.0001169. The correlation coefficient between the noise multiplier and average membership inference attack accuracy was -0.947, with a p-value of 0.00003146. Since the p-values for both hypotheses were significantly lower than the threshold value of 0.05, this provides strong evidence against the null hypotheses. These results are summarized in Table 3, providing evidence to either reject or fail to reject the null hypotheses.

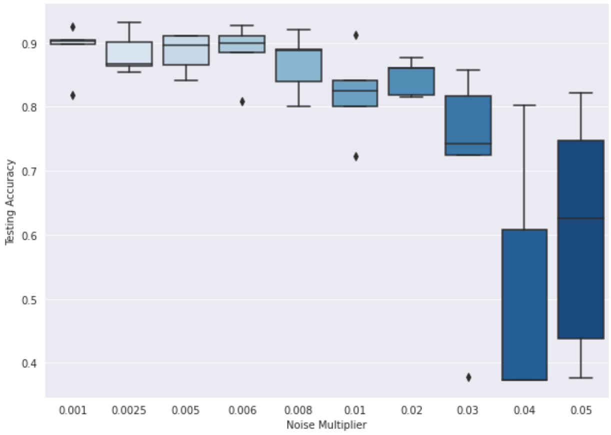

Figure 2: Box plot depicting the relationship between testing accuracy and noise multiplier settings. The figure demonstrates that increasing the noise multiplier leads to a decrease in average testing accuracy.

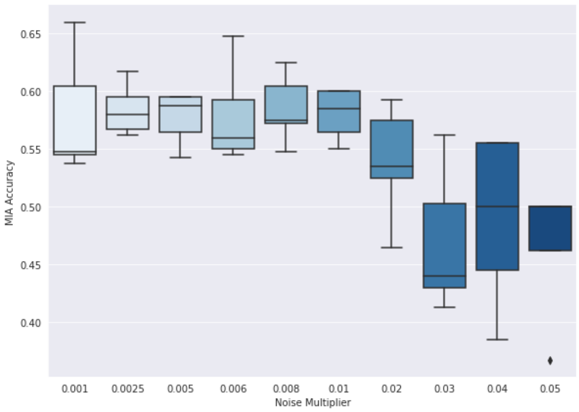

Figure 3: Box plot depicting the relationship between MIA accuracy and noise multiplier settings. The figure demonstrates that increasing the noise multiplier leads to a decrease in average MIA accuracy.

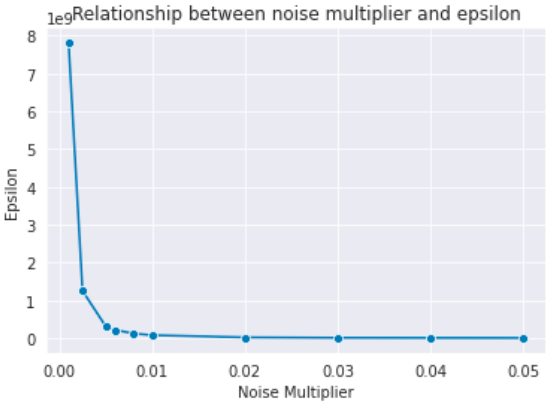

Figure 4: Tradeoff between epsilon (privacy loss metric) and privacy.

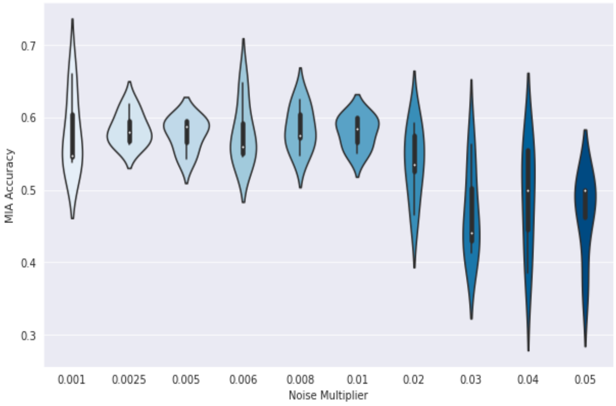

Figure 5: Increased noise level improves privacy and reduces membership attack accuracy, along with amplifying variance in membership inference attacks.

Membership Inference Attack (MIA) Accuracy

Noise Multiplier	*Iteration 1*	*Iteration 2*	*Iteration 3*	*Iteration 4*	*Iteration 5*	*Average MIA Accuracy*
Baseline Model	0.68	0.577	0.6	0.575	0.5775	0.6019
0.001	0.605	0.545	0.66	0.5475	0.5375	0.579
0.0025	0.6175	0.5675	0.58	0.5625	0.595	0.5845
0.005	0.5875	0.565	0.5425	0.595	0.595	0.577
0.006	0.545	0.55	0.56	0.5925	0.6475	0.579
0.008	0.5725	0.605	0.575	0.625	0.5475	0.585
0.01	0.6	0.565	0.585	0.55	0.6	0.58
0.02	0.535	0.575	0.5925	0.465	0.525	0.5385
0.03	0.44	0.5625	0.4125	0.5025	0.43	0.4695
0.04	0.555	0.385	0.445	0.555	0.5	0.488
0.05	0.3675	0.5	0.4625	0.5	0.5	0.466

Table 1: Membership Inference Attack (MIA) Accuracy with Varying Noise Multipliers; As the noise multiplier increases, the average MIA accuracy decreases, indicating enhanced privacy protection

Testing Accuracy

Noise Multiplier	*Iteration 1*	*Iteration 2*	*Iteration 3*	*Iteration 4*	*Iteration 5*	*Average Testing Accuracy*
Baseline Model	0.9021	0.8890	0.8796	0.8844	0.8796	0.8869
0.001	0.8973	0.8973	0.8186	0.9021	0.9021	0.8835
0.0025	0.8668	0.8635	0.8539	0.9005	0.9325	0.8834
0.005	0.8411	0.8652	0.9101	0.9101	0.8956	0.8844
0.006	0.8989	0.9101	0.9278	0.8844	0.809	0.8860
0.008	0.801	0.8876	0.9213	0.8395	0.8892	0.8677
0.01	0.825	0.841	0.801	0.9117	0.7223	0.8202
0.02	0.8764	0.8154	0.8604	0.8604	0.8186	0.8462
0.03	0.817	0.3788	0.7432	0.8571	0.7239	0.7040
0.04	0.374	0.8026	0.6083	0.374	0.374	0.5066
0.05	0.748	0.4382	0.8218	0.626	0.3772	0.6022

Table 2: Testing Accuracy with Varying Noise Multipliers; As the noise multiplier increases, the average testing accuracy decreases, indicating a tradeoff between privacy and accuracy

Null Hypothesis (Ho)	Pearson Correlation Coeff	P-value	Reject Ho
Adding more privacy does not affect model accuracy	-0.9265	0.00011697	✓
Adding more privacy does not affect membership inference attack accuracy	-0.9474	0.00003146	✓

Table 3: Hypothesis testing results for the impact of privacy on model accuracy and membership inference attack accuracy.

Discussion

The experiments conducted on binary classification for pneumonia detection provided evidence supporting the initial hypothesis of a privacy-accuracy and privacy-adversarial attack vulnerability tradeoff. The results indicated that increasing privacy measures, such as employing a noise multiplier, resulted in decreased vulnerability to adversarial attacks. However, it is important to acknowledge that this increased privacy comes at the cost of reduced model accuracy. Therefore, the choice of noise multiplier depends on the tradeoff between the desired level of accuracy and the benefits of privacy and protection against adversarial attacks. Determining the optimal balance between privacy and accuracy is essential and should be based on the specific task requirements, contributing to our understanding of the intricate relationship between privacy and accuracy in machine learning models.

Furthermore, the experimental outcomes validated our hypothesis that differential privacy measures can reduce the risks of membership inference attacks. This highlights the novel approach of considering membership inference attack accuracy as a metric for evaluating privacy. In our study, the baseline models exhibited an MIA accuracy of 0.602, indicating that attackers could successfully guess whether an X-ray image belonged to the training dataset with an accuracy of 60.2%. This relatively high accuracy score is attributed to the white box nature of the attack, where the attack model had information about the architecture of the target model. However, with the introduction of privacy measures, with noise multiplier 0.02, the MIA accuracy was reduced to 0.5385 while maintaining a testing accuracy of 0.8462, making it more challenging for attackers to determine if a sample was part of the training dataset. Note that random chance would imply an MIA accuracy of 0.5. These findings underscore the importance of considering MIA accuracy as a metric to evaluate the privacy achieved by models and the effectiveness of privacy-enhancing techniques.

Future research should explore additional adversarial attacks, such as black box adversarial attacks where the attacker would have limited or no knowledge about the architecture of the target model, or poisoning attacks [24], to validate the robustness of differential privacy guarantees against various attack types. By evaluating the model's susceptibility to black box attacks, researchers can both validate the effectiveness of using membership inference accuracy as a metric and assess the strength of the privacy-enhancing techniques.

Moreover, it was observed that beyond a certain threshold, specifically at high values of the noise multiplier of 0.03 and greater, adjusting the noise multiplier setting led to instability and hindered the model's learning ability. This resulted in significant variance in model performance and MIA accuracy scores. Thus, selecting the appropriate hyperparameter setting requires careful consideration specific to the task at hand. Employing a grid search approach can assist in finding the optimal value for the noise multiplier, ensuring stable model performance. In future studies, exploring other neural network architectures used in medical image analysis tasks, such as U-Net, ResNet, EfficientNet, InceptionNet, and vision transformers, may provide valuable insights into their resilience to privacy attacks.

Furthermore, the applicability of differential privacy extends beyond the binary classification task for pneumonia detection. It can be applied to various medical imaging tasks, including classification, segmentation, and other image analysis tasks. The privacy-accuracy tradeoff observed in this study can guide the development of privacy-preserving models in these domains. However, scaling privacy-preserving techniques to handle larger-scale medical image analysis problems requires consideration of computational challenges and scalability issues. Future research should focus on addressing these challenges and exploring the feasibility of implementing differential privacy in federated learning, where multiple servers collaborate to train a model without sharing sensitive data. By ensuring that individual data contributions remain private, differential privacy enables secure collaboration. Further exploration of the use of differential privacy in federated learning is necessary to ensure advancements in privacy-preserving machine learning approaches.

Limitations

Privacy considerations place limitations on our ability to enhance privacy using techniques like differential privacy. While these measures can improve the privacy of our models and reduce vulnerability to adversarial attacks, there is an inherent trade-off between privacy and model accuracy. Increasing the noise multiplier hyperparameter to large values in order to enhance privacy can significantly impede the model's performance, resulting in notable reductions in accuracy scores.

Another limitation arises from the incompatibility between differential privacy and batch normalization layers. Consequently, when utilizing built-in libraries that implement differential privacy, the substitution of batch normalization becomes necessary. However, alternative normalization methods, such as layer normalization or instance normalization, which do not effectively address the internal covariate shift, must be employed instead.

In addition, it is important to acknowledge that the scope of this study was limited in terms of hyperparameter exploration. Due to time and computational constraints, we focused solely on tweaking the noise multiplier hyperparameter and did not explore the impact of adjusting the gradient clipping threshold. This limitation restricts our ability to fully understand the combined effects of these hyperparameters on model performance and privacy. Future studies should consider investigating the influence of both hyperparameters in order to gain a more comprehensive understanding of their impact on model behavior and the privacy-accuracy tradeoff.

Lastly, it is worth noting that the evaluation of privacy in machine learning models remains a challenging task. Quantifying privacy is inherently difficult, as it requires comprehensive analysis and validation on real-world datasets. In our study, we addressed this challenge by considering membership inference attack accuracy as a metric for evaluating privacy. However, more research is needed to further investigate and refine the methodologies used to assess and quantify privacy in privacy-preserving machine learning approaches.

Conclusion

In conclusion, this study makes a significant contribution to our understanding of the privacy-accuracy tradeoff and the effectiveness of privacy measures against adversarial attacks in the domain of pneumonia detection. The findings underscore the challenging nature of achieving privacy, as it often comes at the expense of model accuracy. Moreover, evaluating the achieved level of privacy presents its own set of difficulties. Future research should aim to address these challenges and develop robust methods for measuring and evaluating privacy in deep learning models. Future research should validate the robustness of differential privacy against additional attack scenarios, explore different CNN architectures in various medical imaging tasks, and extend the study to multiclass classification problems. The insights gained from this study have implications for the implementation of privacy-preserving techniques in healthcare and other sensitive data domains. By considering the tradeoff between privacy and accuracy, differential privacy can be leveraged to ensure privacy in collaborative machine learning settings, such as federated learning, and advance the field of privacy-preserving machine learning.

References

[1] Kaul, D., Raju, H., Tripathy, B.K. (2022). Deep Learning in Healthcare. In: Acharjya, D.P., Mitra, A., Zaman, N. (eds) Deep Learning in Data Analytics. Studies in Big Data, vol 91. Springer, Cham. https://doi.org/10.1007/978-3-030-75855-4_6

[2] Alirr, O. I. (2020, October). Deep learning and level set approach for liver and tumor segmentation from CT scans. Journal of applied clinical medical physics. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7592966/

[3] Panfilov E;Tiulpin A;Nieminen MT;Saarakkala S;Casula V; (n.d.). Deep learning-based segmentation of knee MRI for fully automatic subregional morphological assessment of cartilage tissues: Data from the osteoarthritis initiative. Journal of orthopaedic research : official publication of the Orthopaedic Research Society. https://pubmed.ncbi.nlm.nih.gov/34324223/

[4] Chen, I. Y., Pierson, E., Rose, S., Joshi, S., Ferryman, K., &amp; Ghassemi, M. (2021, July). Ethical machine learning in Healthcare. Annual review of biomedical data science. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8362902/

[5] (OCR), O. for C. R. (2023, April 30). Hipaa Home. HHS.gov. https://www.hhs.gov/hipaa/index.html

[6] Bluemke, E. (2021, November 17). Privacy-preserving AI in medical imaging: Federated Learning, Differential Privacy, and encrypted computation. OpenMined Blog. https://blog.openmined.org/federated-learning-differential-privacy-and-encrypted-computation-for-medical-imaging/

[7] Jain, P., Gyanchandani, M., &amp; Khare, N. (2018, April 13). Differential Privacy: Its technological prescriptive using big data - journal of big data. SpringerOpen. https://journalofbigdata.springeropen.com/articles/10.1186/s40537-018-0124-9

[8] Kamran, S., Munir, A., Raza, B., Ahmad, J., & Baik, S. W. (2020). Evaluation of Inference Attack Models for Deep Learning on Medical Data. ResearchGate. https://www.researchgate.net/publication/345215882_Evaluation_of_Inference_Attack_Models_for_Deep_Learning_on_Medical_Data

[9] Hongsheng Hu The University of Auckland, Hu, H., Auckland, T. U. of, Zoran Salcic The University of Auckland, Salcic, Z., University, L. S. L., Sun, L., University, L., Gillian Dobbie The University of Auckland, Dobbie, G., Philip S. Yu University of Illinois at Chicago, Yu, P. S., Chicago, U. of I. at, University, X. Z. M., Zhang, X., University, M., Martin, A., Mohanad, A., Martin, A., … Metrics, O. M. A. (2022, January 1). Membership inference attacks on Machine Learning: A Survey. ACM Computing Surveys. https://dl.acm.org/doi/10.1145/3523273

[10] Hao, S., Zhang, Y., Yang, X., & Jin, R. (2021). On the Privacy Risks of Adaptive Federated Learning. arXiv preprint arXiv:2105.02866

[11] NIST. (2021, March 18). Differential Privacy: A Privacy-Preserving Data Analysis. Retrieved from https://www.nist.gov/blogs/cybersecurity-insights/differential-privacy-privacy-preserving-data-analysis-introduction-our

[12] Huang, R. (2021, June 17). Understanding Differential Privacy. Towards Data Science. Retrieved from https://towardsdatascience.com/understanding-differential-privacy-85ce191e198a

[13] Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K., & Zhang, L. (2020). Deep learning with differential privacy. arXiv preprint arXiv:2006.03572.

[14] LIN, Ying; BAO, Ling-Yan; LI, Ze-Minghui; SI, Shu-Sheng; and CHU, Chao-Hsien. Differential privacy protection over deep learning: An investigation of its impacted factors. (2020). Computers & Security. 99, 1-16. Research Collection School Of Computing and Information Systems.

[15] Yu, D., Kamath, G., Kulkarni, J., Liu, T.-Y., Yin, J., & Zhang, H. (2023, February 5). Individual privacy accounting for differentially private stochastic gradient descent. arXiv.org. https://arxiv.org/abs/2206.02617

[16] Inference attacks against Machine Learning Models. Predict the future. (2020, September 6). https://techairesearch.com/inference-attacks-against-machine-learning-models/

[17] Chen, Y., Shen, C., Shen, Y., Wang, C., &amp; Zhang, Y. (2022, May 16). Amplifying membership exposure via data poisoning. OpenReview. https://openreview.net/forum?id=mT18WLu9J

[18] Kermany, D. (2018, January 6). Labeled optical coherence tomography (OCT) and chest X-ray images for classification. Mendeley Data. https://data.mendeley.com/datasets/rscbjbr9sj/2

[19] Mooney, P. (2018, March 24). Chest X-ray images (pneumonia). Kaggle. https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia

[20] Handlin, C. W. (2022, March 2). Reproducible data science and why it matters. Medium. https://engineering.rappi.com/reproducible-data-science-and-why-it-matters-e4e62fd60b9a

[21] Zhou, T., Ye, X., Lu, H., Zheng, X., Qiu, S., &amp; Liu, Y. (2022, April 25). Dense convolutional network and its application in medical image analysis. BioMed Research International. https://www.hindawi.com/journals/bmri/2022/2384830/

[22] Spring-Epfl. (n.d.). Spring-EPFL/MIA: A library for running membership inference attacks against ML Models. GitHub. https://github.com/spring-epfl/mia

[23] Shokri, R., Stronati, M., Song, C., & Shmatikov, V. (2017, March 31). Membership inference attacks against Machine Learning Models. arXiv.org. https://arxiv.org/abs/1610.05820

[24] Lin, J., Dang, L., Rahouti, M., & Xiong, K. (2021, December 6). *ML attack models: Adversarial attacks and data poisoning attacks*. arXiv.org. https://arxiv.org/abs/2112.02797