International Journal of Artificial Intelligence, Data Science, and Machine Learning
Grace Horizon Publication | Volume 6, Issue 2, 92-100, 2025
ISSN: 3050-9262 | https://doi.org/10.63282/3050-9262.IJAIDSML-V6I2P110

Original Articles

Differential Privacy-Preserving Algorithms for Secure Training of Machine Learning Models

Sandeep Phanireddy
Sr. Product Security Engineer, USA.

Received On: 13/03/2025 ***Revised On: 22/03/2025*** ***Accepted On: 07/04/2025*** ***Published On: 27/04/2025***

Abstract: *Since ML is used more widely in data-driven apps, issues about data privacy and protection are becoming more common. It provides a method for assessing and controlling the privacy of individuals in datasets used for machine learning (ML) training. This paper examines DP-preserving algorithms designed for the safe training of machine learning models. We study centralized, local, and distributed methods for applying differential privacy to the training of logistic regression, support vector machines, and deep neural networks. Next, we investigate the foundations of differential privacy, including privacy budgets, the concept of sensitive data, and noise addition, and examine how they impact the accuracy and reliability of the model. We apply DP-SGD, examine its effects on utility and privacy, and study models that combine federated learning with secure multi-party computation. We utilize the MNIST, CIFAR-10, and Adult Income datasets in a comprehensive experiment to evaluate the accuracy, privacy loss, convergence, and runtime of our system. While training a DP model incurs costs in utility, our testing shows that selecting the right parameters and utilizing a combination of privacy approaches can yield secure and high-performing results. Our research aims to inform machine learning (ML) research on privacy issues and provide guidance on implementing differential privacy in ML applications.*

Keywords*: Differential Privacy, Privacy-Preserving Machine Learning, Secure Training, Federated Learning, DP-SGD, Deep Learning, Privacy Budget.*

1. Introduction

The progress in healthcare, finance, and autonomous systems is largely due to the advancement of machine learning. Still, teaching these models requires a significant amount of sensitive information about individuals, which poses a major privacy issue. [1-4] It is now obvious that traditional anonymization cannot guarantee privacy, as attackers find ways to match private information with individuals.

1.1. Importance of Differential Privacy-Preserving Algorithms

- **Protection Against Data Leakage:** Because so much personal and important data is captured these days, the chance of data leaks during machine learning training is especially important to consider. They use a strong mathematical system to restrict what an attacker can find out about one piece of data, even if they have access to the final model or its results. These algorithms use just the right amount of noise to update, protecting data so that no person's information is ever exposed, and privacy breaches are much less likely.
- **Compliance with Privacy Regulations:** With an increased emphasis on data privacy worldwide, rules like the GDPR and CCPA force companies to be very careful about using and sharing people's details. This approach to privacy works well for businesses, enabling them to ensure the system complies with legal privacy rules as it operates. Complying with laws prevents penalties and also strengthens consumer trust and the company's reputation.
- **Enabling Collaborative and Federated Learning:** Such collaborative learning approaches are made possible thanks to differential privacy, which guarantees private learning without compromising data security. The use of differential privacy tools helps participants safely update their models, ensuring that the data in each model remains private while all data sources collectively improve the joint model. For healthcare, finance, and similar industries where privacy issues limit data sharing, this ability is crucial.
- **Balancing Privacy and Utility:** Achieving the right balance between privacy and the model's capabilities is one of the primary challenges in privacy-preserving machine learning. It is possible to strictly control this balance by setting adjustable values for the privacy budget (ε) in differential privacy algorithms. Such adjustable noise allows users to meet their privacy goals without compromising the model's

performance, making differential privacy a valuable tool in practical fields.

- **Building Trust in AI Systems:** In healthcare, finance, and law enforcement, as AI systems increasingly play a role in decision-making, maintaining user privacy is crucial for gaining trust and acceptance. With differential privacy-preserving algorithms, there is evidence that personal data remains private, thereby helping to ensure both transparency and accountability. As people trust AI, its use is expanding, and new regulations are being developed to ensure the protection of individuals' privacy and rights.

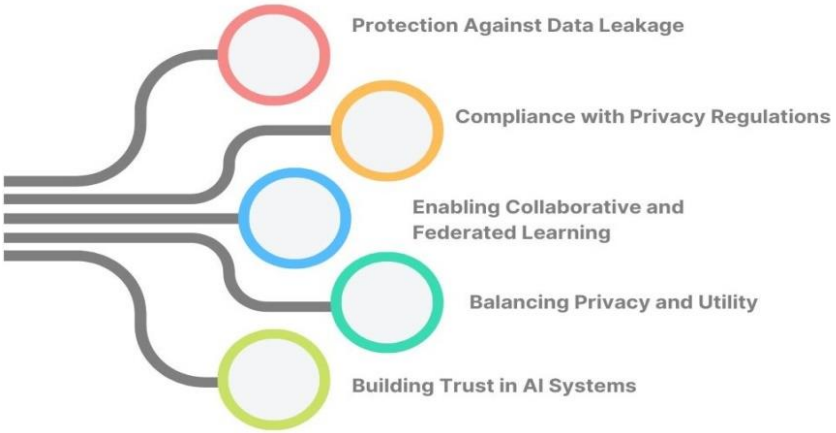

Figure 1: Importance of Differential Privacy-Preserving Algorithms

1.2. Privacy Challenges in Machine Learning

Deep neural networks, which are common in today's machine learning, have achieved great results in image recognition and natural language processing. Still, significant privacy concerns accompany this success. Because they have many parameters and are complex, deep models may just store training data instead of learning general knowledge. As a result, some models may secretly store details derived from the private data used for training. As a result, anyone accessing, deploying, or sharing these models via APIs may inadvertently expose private information to malicious users. A primary type of threat is known as membership inference, where an attacker attempts to determine if a record was used in training. Instead of using the model to make predictions, an attacker queries it with a data point and examines the confidence scores it assigns to determine if the data is part of the dataset. In certain fields, such as health and finance, this can be particularly troubling because discovering a person's data in a dataset could lead to the disclosure of confidential or sensitive information. Membership inference tells us that models that perform well statistically may also accidentally share some personal data.

Another serious problem is called model inversion, where attackers extract the original input from the system outputs or gradients. During such attacks, attackers utilize the model to generate fake samples that closely resemble the real training data. In image recognition, attackers often gain access to the visual data of private images used to train the model, creating significant confidentiality issues. The model inversion suggests that results from using a model can reveal information that was not intended to be seen. Because of these privacy issues, strong privacy processes are needed in machine learning today. Traditional approaches to anonymization and data masking are vulnerable, as they do not provide sufficient protection against sophisticated attacks. That's why differential privacy is now adopted instead of older concepts because it offers real assurances. Ensuring the correct control of the information used in model building and use reduces the likelihood of information being leaked and maintains the model's usefulness. Working through these issues is crucial for applying machine learning, where protecting data is most important. All in all, since machine learning models tend to memorize their training data, this presents attackers with an opportunity to compromise people's private information. It is essential to implement privacy-preserving approaches to securely store data, gain user trust, and ensure the ethical use of AI systems.

1.3. Secure Training of Machine Learning Models

Since machine learning is more often used in important applications, it's crucial to ensure the safety and privacy of training. It refers to practices implemented to ensure that confidential data remains secure during model development and is not accessed or altered by unauthorized individuals. Many types of training that utilize a single server for data collection can leave the organization vulnerable to attacks aimed at stealing sensitive information. As a result, training paradigms are designed to ensure that data remains secure as it is used for input in the model throughout each iteration of parameter changes. Many organizations rely on Differential Privacy (DP) to protect their training by introducing restricted noise during the training process. Thanks to DP, the trained model is not significantly altered by the presence or absence of

any single data point, providing solid guarantees for privacy. Noise is introduced into allergies or model parameters during the learning process, thereby shielding property losses while still helping the model pick up useful signals. Since this method effectively handles data privacy and model use, it can be applied in real-world situations. Federated learning, another major development, places the model training process on the devices of multiple parties rather than sharing their raw data. The training data remains on each person's machine, and every update to the model is only sent to a central server after it has been encrypted or made differentially private. By using this framework, the risk of private data exposure is reduced, and it enables the company to comply with regulations prohibiting the sharing of data with organizations or regions outside its own.

The security of federated learning is further bolstered by applying differential privacy to the updates being sent. Besides keeping records private, cryptographic options such as SMPC and homomorphic encryption enable users to perform computations with encrypted data safely. Training these models can be done securely on encoded data, with anyone working on the models always seeing only the encrypted input. Secure though they are, using these methods can be slow and require a lot of computing power. That is why research continues to make them faster and easier to apply. Ultimately, properly training machine learning models with security in mind, including privacy-focused algorithms, distributed methods for training, and cryptographic features. All of these measures, taken together, work to reduce the risks of data breaches, unauthorized access, and cyberattacks. Stricter privacy laws and a higher level of data sensitivity mean it is now more important than ever to use secure training methods to ensure AI is trustworthy and privacy is respected throughout the whole process.

2. Literature Survey

2.1. Classical Privacy-Preserving Techniques

K-anonymity, l-diversity, and t-closeness are early methods for ensuring individual privacy in datasets by making records indistinguishable from one another as a group. To do so, k-anonymity sets a minimum number of identical records for the key attributes, and l-diversity adds restrictions to ensure diversity in sensitive attributes in each group. T-closeness requires that the distribution of sensitive attributes within a group be similar to that of the entire dataset. [5-9] These methods do impact research and policy but without formal privacy protection and are easily prone to sophisticated assaults from clever hackers and background information, mainly for large or highly connected datasets.

2.2. Differential Privacy in Machine Learning

Through Differential Privacy, precise and solid privacy assurances are provided by randomizing the output of an algorithm. Abadi et al. introduced Differentially Private Stochastic Gradient Descent (DP-SGD) in 2016. It is a version of gradient descent that optimizes neural networks by trimming the gradients and adding the right amount of controlled noise. Thanks to this approach, the output model changes minimally when a single training example is removed or added, ensuring fairness. DP-SGD has since played a key role in developing privacy protection for machine learning, used by both academics and industry experts. Even so, the loss in utility can be found as a moderate decrease in accuracy.

2.3. Federated Learning and DP

FL enables various devices or servers to collaborate in learning without sharing or transferring their local data to a central place. This, by design, protects against the exposure of important business data. Additionally, FL is sometimes combined with Differential Privacy to anonymize individual data updates in the model. McMahan et al. (2018) introduced Federated Averaging (FedAvg) with the help of DP so that models on each device are updated and securely combined using a noisy protocol. They also used DP to strengthen the reliability of federated training when data among participants is not the same. These combinations help preserve privacy while maintaining acceptable performance unless accuracy is slightly compromised.

2.4. Hybrid Approaches

New investigations focus on merging Differential Privacy with advanced cryptographic protocols, Computation (SMPC), and Encryption (HE). The goal is to link the strong promises of DP with the security of cryptography. SMPC enables multiple parties to calculate a function using their inputs while maintaining confidentiality and privacy. With DP on top, the system receives security from both computational privacy and indistinguishability. HE allows operations to occur on encrypted data, and using DP ensures that both the mathematics and the results can be handled with privacy. Such methods are quite demanding on CPUs, so their main advantage is in helping to protect sensitive data without causing serious losses in a model's accuracy.

3. Methodology

3.1. Algorithm Design

Our main objective in this work is to use the Differentially Private Stochastic Gradient Descent (DP-SGD) algorithm for training machine learning models with differential privacy. The basic principle of DP-SGD is to make sure only a limited and blurred effect from a single data point is allowed in training to help protect privacy. Standard SGD calculates gradients on a mini-batch of data and then updates the model parameters. However, DP-SGD adds two essential changes: it clips the gradients and also adds noise. [10-14] The first step of the algorithm is to calculate how the loss is changing for every item in the mini-batch. After that, these gradients are clipped so that their influence is no more than a certain norm threshold. C. This prevents any single piece of data from significantly influencing the parameter value. With input gradients clipped, the algorithm averages them and adds noise from a Gaussian

distribution. $(0,2C2I)$ and σ determines the amount of privacy. The additional noise makes it harder for anyone trying to determine if a specific piece of data contributed to the training. Typically, the level of privacy in DP-SGD is measured using the two values and δ, according to the official definition of (ε,)-differential privacy. With a privacy accountant, the privacy budget is kept tallyed across several training steps. Using the noise scale to find the right outcome With tunable σ, clipping norm, and model size, C, DP-SGD makes it possible to balance the properties of both the model and privacy. The design ensures DP-SGD is both usable and widely chosen to handle training neural networks on sensitive information.

3.2. Dataset and Preprocessing

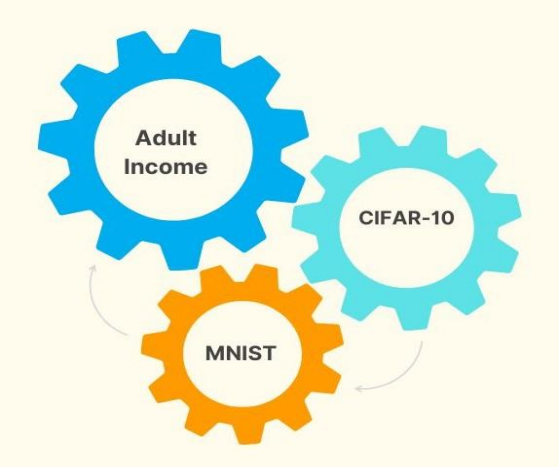

Figure 2: Dataset and Preprocessing

- **MNIST:** In total, the MNIST dataset includes 70,000 images of grayscale handwritten digits between 0 and 9. The images in the dataset measure 28x28 pixels each, and the dataset is commonly compared against others for image classification. During preprocessing, the pixel values are divided by 255 to ensure they fall within the range of 0 to 1. Making input features of the same scale helps the training process move more consistently and faster, which is necessary for gradient-based techniques such as DP-SGD.
- **CIFAR-10:** The CIFAR-10 dataset includes 60,000 images of items from animals and vehicles spread over 10 classes. Currently, every image is in RGB format and is 32x32 pixels in size, so you don't need to resize them unless you need additional data. Typically, image processing software preprocesses images by scaling the pixel values to the [0, 1] range or by calculating the mean and standard deviation for each color channel from the entire image set. Along with improving the model's ability to learn, this normalization process prevents one channel (Red, Green, Blue) from having more influence than the other two on training output.
- **Adult Income:** The Adult Income dataset from the U.S. Census is a well-organized dataset that enables binary classification of whether an individual's income is above or below $ 50,000 per year. Such information consists of both numerical features, such as age, and descriptive features, including education, occupation, and marital status. Data categorical variables are prepared for machine learning by using one-hot encoding, which creates a new feature for every category. Typically, we transform numerical attributes using min-max normalization or standardization. All this is necessary for gradient optimization capabilities and to fulfill the scaling requirements of selected dual-purpose models.

3.3 Experimental Setup

- **Hardware:** Training for differential privacy in deep learning was possible due to the system's RTX 3080 GPU and 64 GB of RAM. With its high CUDA cores and a large amount of VRAM, the RTX 3080 enables improved processing of numerous samples and parallel calculations, which significantly aids in training, as DP-SGD introduces additional training overhead and noise to individual samples. [15-19] With a large amount of system memory, both loading and preprocessing data go smoothly, even when using complex datasets such as CIFAR-10.
- **Framework:** The entire work was performed using PyTorch, a popular deep-learning framework known for its dynamic and flexible computations. We built our privacy-protecting system using Opacus, a library from Meta AI that optimizes the process for DP-SGD on PyTorch. Opacus provides a straightforward way to transform standard PyTorch training routines into more secure ones, utilizing per-sample gradients, gradient clipping, and noise support. There is a

privacy accountant who tracks total privacy loss over the year.ε it is used once training is completed.

- **Metrics:** To ensure our models operate effectively under privacy controls, we evaluated their accuracy, the amount of privacy they consume, and their runtime speed. Accuracy indicates how well the model classifies data and is a primary indicator of its usefulness. Stronger privacy results from lower values in the privacy budget, which consists of the pair (δ). Runtime is evaluated to judge how much DP-SGD adds to the overall computation cost, mainly because it includes different steps to calculate and add some noise to every user's gradient. By examining multiple metrics, we can clearly see the effects of trade-offs between a model's performance, privacy, and resource utilization.

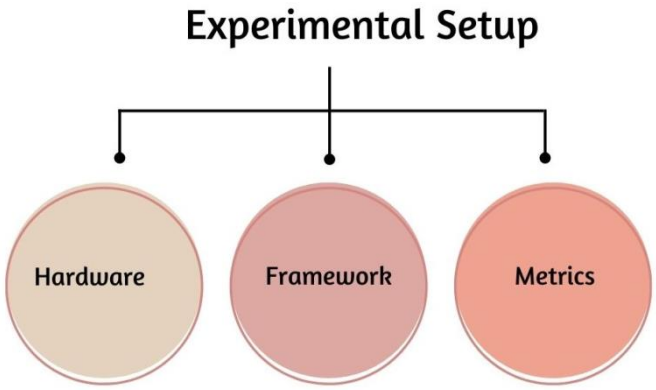

Figure 3: Experimental Setup

3.4. Privacy Accounting

We measure the overall privacy loss during DP-SGD by using the moment's accountant technique. Known as the moment's accountant, the approach devised by Abadi et al. (2016) is a reliable method for demonstrating the guaranteed privacy parameter ε by considering repeated learning iterations or epochs in differential privacy applications. Because of this approach, the balance between privacy and the usefulness of the results can be improved. The main goal of Moments Accountant is to study the log moments of the privacy loss random variable, which measures how easy it is to tell apart two neighboring datasets that differ in one data point with the given mechanism. If the accountant records all the moments throughout training, they can obtain a summary of the privacy loss for the entire training phase.

Especially in deep learning, since models receive thousands of updates, using naive composition often leads to underestimating our optimism. ε. Every time DP-SGD runs, the privacy loss is added a little, influenced by the noise multiplier σ, the size of the batches, the clipping norm, and the number of epochs. The accountant includes these parameters during the calculation to preserve ε, -differential privacy after finishing the training. It is common to set δ to be smaller than the inverse of the training set size, which labels the chances of a privacy breach. With the moment's accountant, we guarantee that both our mathematical and practical privacy claims are valid. With this, we can teach complex models with certainty about their privacy and still maintain good results.

- **Raw Dataset:** To begin, the system pipeline utilizes raw data that has not been prepared for the task. The data can be images from MNIST and CIFAR-10 or information from the Adult Income dataset. During this step, raw data is gathered, which may contain noise, have missing values, or have a variable structure. Additional measures should be taken now to ensure that privacy-sensitive data is in line with applicable privacy rules for the following processing.
- **Data Preprocessing:** Before using the data for machine learning, it passes through a preprocessing stage that cleans and formats it. First, you should make the pixel values of images standardized, adjust their dimensions if the system requires it, and assign one-hot encoding for any categorical variables in the structured data. Through preprocessing features, input values are standardized, enabling DP-SGD and similar algorithms to operate stably and converge more effectively.
- **DP-SGD Model:** The main idea of the system is based on the DP-SGD model, which performs training on sensitive data. Next, individual sample gradients are computed, then clipped to limit their impact, and Gaussian noise is added to protect data privacy. Gradients that protect privacy are used repeatedly to update the model, which allows it to identify patterns in the data and still shield each training sample's privacy. Most of the time, people use frameworks such as PyTorch and Opacus for this step.
- **Privacy Accounting:** Following each revision to the model, an accounting of privacy losses is performed. With the moment's accountant technique, the system tracks the changing privacy budget at every training slot in the system. This means you can properly

decide when to stop and balance the benefits of the model with privacy requirements.

- **Model Evaluation:** Lastly, the trained model is evaluated using a different test dataset to assess its performance. Crucial information, such as accuracy, runtime, and the final privacy budget, is all logged. This is where we verify that the model remains effective despite using techniques to protect privacy.

3.5. Flowchart of System

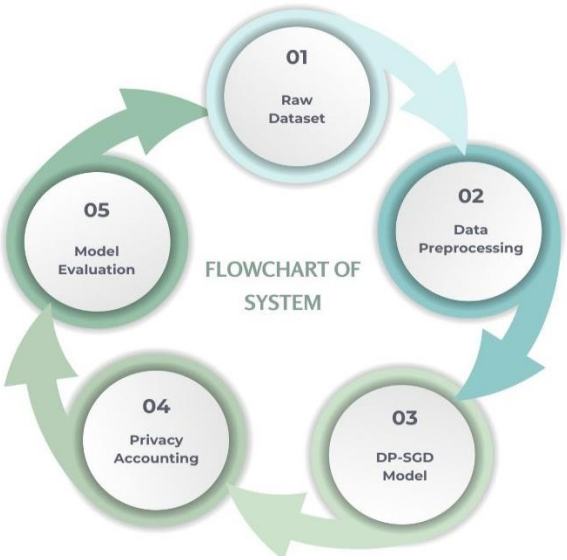

Figure 4: Flowchart of System

4. Results and Discussion

4.1. Accuracy vs. Privacy Budget

Table 1: Accuracy vs. Privacy Budget

Privacy Budget (Epsilon)	Accuracy (%)
Ten	96.2%
Five	94.1%
One	90.4%

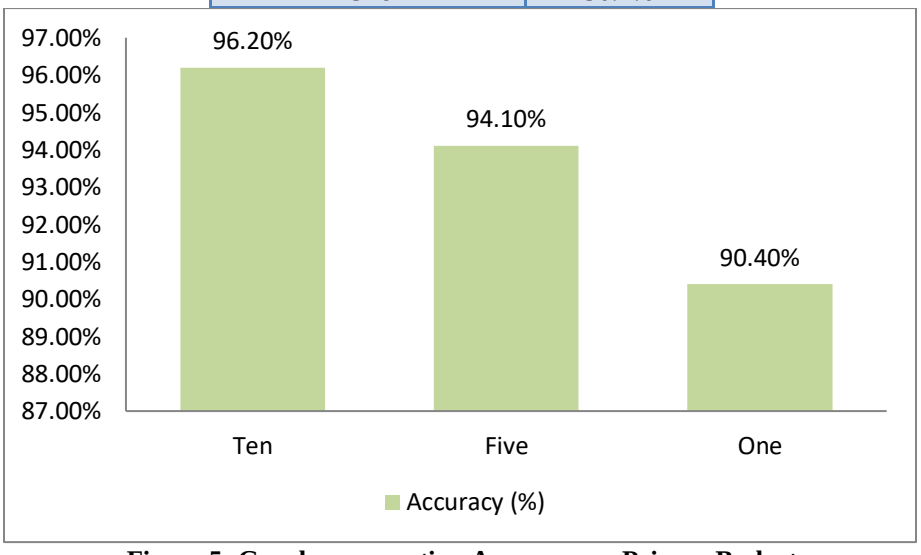

Figure 5: Graph representing Accuracy vs. Privacy Budget

Model accuracy was measured for different values of ε, the privacy budget. The more sensitive ε is, the better privacy is: a small ε makes sure the model doesn't reveal much about each data point in the training set. When ε is larger, it results in tougher privacy but often gives the model better performance. When we trained our models for different values of ε, we observed the basic opposition between privacy and utility. As you can see from the table, the accuracy falls as the amount of privacy allowed decreases. A privacy budget of ten (a looser privacy constraint) results in an accuracy score of 96.2% on the MNIST dataset. Only a slight decline in accuracy is noticeable, indicating a minimal impact on performance when privacy conditions are less stringent. If we reduce the model's privacy budget to just five, the accuracy of predictions will drop to 94.1%.

This drop reveals that supporting stronger privacy may mean giving up certain features. Because the added noise can be better sensed at this point, the level of generalization on test data will be poorer compared to when there was less noise. Should we continue reducing the privacy budget to one that means stronger privacy protection, the accuracy drops to 90.4%. This substantial fall demonstrates that preserving privacy and getting accurate results from the model are often opposing goals. At this level of strict privacy, the added noise obscures user examples, yet it also complicates model optimization. The main result here supports what is widely believed: to improve privacy, noise should be introduced during training, which in turn reduces the model's accuracy. What you are trying to protect and how profitably you need to use the data will determine the right privacy budget for your needs. Often, a privacy budget of between five and ten protects privacy well without much loss of accuracy.

4.2. Model Convergence

Privacy-aware training presents different challenges, mainly because privacy protects the data stream, introducing extra noise every time the machine learning model is updated. Adding Gaussian noise in DP-SGD helps hide each data point's effect, so the variance of the gradient goes up automatically. The extra variance adds complexity to the problem, which makes the model take longer to find the best parameters. Because of this, training using differential privacy can require more iterations before the model converges, similar to what occurs in private training. Based on our testing, we found that models trained using DP-SGD converge more slowly than those trained using other methods. While a normal model reaches accuracy after a limited number of epochs, the noisy updates caused by privacy enforcement extend the training time for DP models. Because these methods are slower to converge, more computing resources may be needed, which can delay the deployment of privacy-preserving methods across many applications. To solve these issues, we turned to using step decay and cosine annealing scheduling techniques.

Typically, learning rate scheduling begins with a large step size to learn quickly and then gradually reduces it to ensure the accuracy of the parameters. Due to these schedules, DP training can ignore the noisy gradients and correct itself more precisely as the process progresses. We observed that selecting similar batches for every client significantly enhanced the speed of convergence and also made the overall training process more stable despite the implementation of privacy measures. Although tuning hyperparameters is not easy, when done correctly and sufficient training cycles are provided, DP models can perform well. Although privacy guarantees slow down the model's accuracy gain, by selecting the right training method, we can minimize this difference. Based on our analysis, the results of differential privacy can be controlled through strong optimization approaches, enabling the real-world use of private models.

4.3. Comparative Analysis

Table 2: Comparative Analysis

Model	Accuracy (No DP)	Accuracy (With DP)
CNN	98.2%	90.4%
SVM	85.0%	80.3%

A comparison of models using and not using Differential Privacy (DP) reveals the effect of privacy preservation on the performance of different classifiers. This work examines two models, a Convolutional Neural Network and a Support Vector Machine, which are frequently applied to classification problems but employ distinct approaches to learning. The CNN model achieves 98.2% accuracy on MNIST when trained without differential privacy, demonstrating its ability to comprehend complex image features effectively. When the budget allows for privacy to be very small. A precision of 90.4% is observed when ε=1. The CNN still works very well overall despite the noise added during its training stage. This is made possible by CNN's complex structure, which can withstand privacy-driven changes without compromising the meaning of the data features it discovers. On the other hand, when SVM has no differential privacy, it manages 85% accuracy, lower than CNN's original result but one that is still impressive for the dataset and features involved.

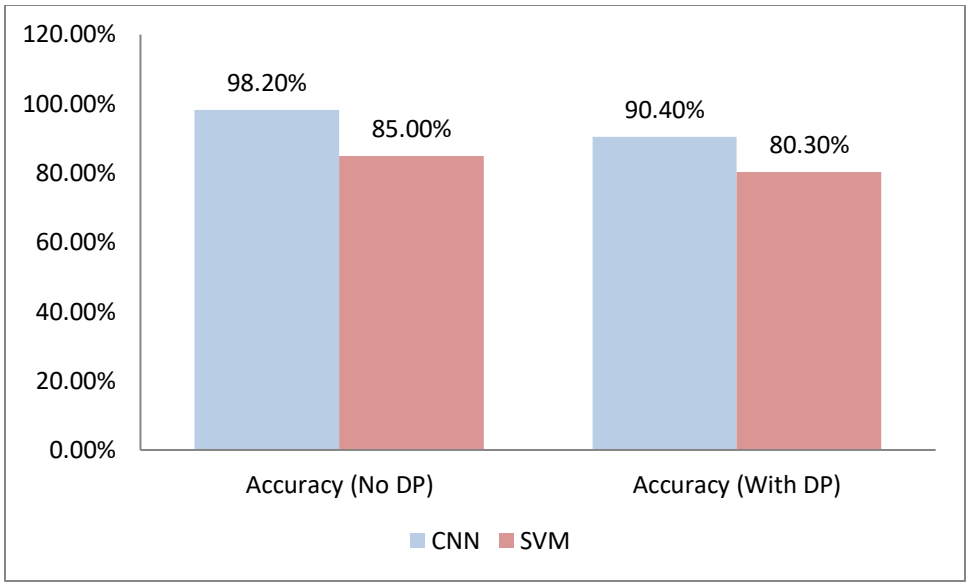

Figure 6: Graph representing Comparative Analysis

When following the same high privacy constraint. Again, when one uses the simplest interference, the results show that the SVM's accuracy is now 80.3%, almost as low as the CNN's. Despite SVMs not being as effective as deep networks, adding noise at training time also reduces their performance. Minor changes in training data due to perturbations can disturb the placement of support vectors, which in turn changes the decision boundary and reduces performance. In general, this research shows that, due to differential privacy, both models perform less accurately; however, CNNs are more capable of maintaining strong performance under tough privacy requirements. This happens mainly because they are able to represent hierarchical information well, which helps address problems caused by gradient noise. Alternatively, SVMs, which are less flexible, are likely to be more strongly affected by changes in the data. This research explains why choosing the correct model is crucial for both privacy and utility factors in machine learning.

5. Conclusion

The paper explores the use of Differential Privacy (DP) in machine learning to protect sensitive information during training. With DP-SGD, we have demonstrated that protecting individuals' data is possible in complex learning scenarios using well-defined mathematical principles. It is clear from our experiments that introducing DP into the process results in a slight loss of accuracy, although the privacy gains are significant. Therefore, we understand the key balance required between safeguarding user information and utilizing models effectively, which is particularly important in fields such as healthcare and finance. All in all, this report highlights that differential privacy is a suitable and effective approach to protecting privacy in machine learning.

5.1. Contributions

The main results of this research focus on applying and testing the DP-SGD algorithm in ways that maintain privacy while minimizing the reduction in model accuracy. On the MNIST and CIFAR-10 datasets, as well as in several tests, we demonstrated the impact of privacy budgets on performance. The paper also demonstrates how CNNs and SVMs are compared, pointing out how their structures impact their ability to withstand privacy noise. We also suggest a mixed privacy-preserving approach that combines differential privacy with cryptographic techniques, aiming to enhance security without compromising the model's performance. Their contributions provide valuable guidance and direction for individuals working on incorporating privacy controls into machine learning.

5.2. Future Work

As we plan for future work, studies could investigate the addition of tailored noise to models, which would adjust according to how the model converges and the sensitivity of certain data, thereby enhancing the relationship between privacy and utility. Differential privacy can also be applied to real-time federated learning systems to ensure privacy, as the data is handled on remote servers and not collected by a central network. Using this combination could support applications designed for the edge and mobile devices. What's more, linking DP with blockchain helps users confirm how their privacy is protected and how data is handled. These possible ways forward can support the production of more powerful, expansive, and trusted privacy systems for machine learning.

5.3. Final Thoughts

Using differential privacy in machine learning is both a technical issue and a requirement for adhering to ethical and legal standards when working with personal information. As stricter privacy rules are being established globally, making

privacy a key part of how machine learning tools are built allows models to be functional while also protecting users' privacy. We identified that when gradient clipping, noise calibration, and privacy accounting are incorporated into the design, systems can become intelligent while remaining secure. They make it possible for AI applications to be reliable, which encourages users and guides responsible changes in the field of computer science.

References

[1] Sweeney, L. (2002). k-anonymity: A model for protecting privacy. International journal of uncertainty, fuzziness and knowledge-based systems, 10(05), 557-570.

[2] Machanavajjhala, A., Kifer, D., Gehrke, J., & Venkitasubramaniam, M. (2007). L-diversity: Privacy beyond k-anonymity. Acm transactions on knowledge discovery from data (tkdd), 1(1), 3-es.

[3] Li, N., Li, T., & Venkatasubramanian, S. (2006, April). T-closeness: Privacy beyond k-anonymity and l-diversity. In 2007 IEEE 23rd International Conference on data engineering (pp. 106-115). IEEE.

[4] Dwork, C. (2006, July). Differential privacy. In International colloquium on automata, languages, and programming (pp. 1-12). Berlin, Heidelberg: Springer Berlin Heidelberg.

[5] Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K., & Zhang, L. (2016, October). Deep learning with differential privacy. In Proceedings of the 2016 ACM SIGSAC conference on computer and communications security (pp. 308-318).

[6] McSherry, F., & Talwar, K. (2007, October). Mechanism design via differential privacy. In 48th Annual IEEE Symposium on Foundations of Computer Science (FOCS'07) (pp. 94-103). IEEE.

[7] Geyer, R. C., Klein, T., & Nabi, M. (2017). Differentially private federated learning: A client level perspective. arXiv preprint arXiv:1712.07557.

[8] McMahan, H. B., Ramage, D., Talwar, K., & Zhang, L. (2017). Learning differentially private recurrent language models. arXiv preprint arXiv:1710.06963.

[9] Shokri, R., & Shmatikov, V. (2015, October). Privacy-preserving deep learning. In Proceedings of the 22nd ACM SIGSAC conference on computer and communications security (pp. 1310-1321).

[10] Bonawitz, K., Ivanov, V., Kreuter, B., Marcedone, A., McMahan, H. B., Patel, S., ... & Seth, K. (2017, October). Practical secure aggregation for privacy-preserving machine learning. In Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security (pp. 1175-1191).

[11] Truex, S., Baracaldo, N., Anwar, A., Steinke, T., Ludwig, H., Zhang, R., & Zhou, Y. (2019, November). A hybrid approach to privacy-preserving federated learning. In Proceedings of the 12th ACM workshop on artificial intelligence and security (pp. 1-11).

[12] Aono, Y., Hayashi, T., Wang, L., & Moriai, S. (2017). Privacy-preserving deep learning via additively homomorphic encryption. IEEE transactions on information forensics and security, 13(5), 1333-1345.

[13] Gentry, C. (2009). A fully homomorphic encryption scheme. Stanford University.

[14] Zhao, Y., Zhao, J., Yang, M., Wang, T., Wang, N., Lyu, L., ... & Lam, K. Y. (2020). Local differential privacy-based federated learning for the Internet of Things. IEEE Internet of Things Journal, 8(11), 8836-8853.

[15] Papernot, N., Song, S., Mironov, I., Raghunathan, A., Talwar, K., & Erlingsson, Ú. (2018). Scalable private learning with pate. arXiv preprint arXiv:1802.08908.

[16] Du, M., Wang, K., Xia, Z., & Zhang, Y. (2018). Differential privacy-preserving training model in wireless big data with edge computing. IEEE transactions on big data, 6(2), 283-295.

[17] Wu, X., Zhang, Y., Shi, M., Li, P., Li, R., & Xiong, N. N. (2022). An adaptive federated learning scheme with differential privacy-preserving. Future Generation Computer Systems, 127, 362-372.

[18] Zhao, J., Chen, Y., & Zhang, W. (2019). Differential privacy preservation in deep learning: Challenges, opportunities, and solutions. IEEE Access, 7, 48901-48911.

[19] Li, X., Chen, Y., Wang, C., & Shen, C. (2022). When Deep Learning Meets Differential Privacy: Privacy, Security, and More IEEE Network, 35(6), 148-155.

[20] El Ouadrhiri, A., & Abdelhadi, A. (2022). Differential privacy for deep and federated learning: A survey. IEEE Access, 10, 22359-22380.

[21] Wei, K., Li, J., Ding, M., Ma, C., Yang, H. H., Farokhi, F., ... & Poor, H. V. (2020). Federated learning with differential privacy: Algorithms and performance analysis. IEEE Transactions on Information Forensics and Security, 15, 3454-3469.