Exploring Strategies for Privacy-Preserving Machine Learning in Distributed Environments

Suresh Dodda[1], Anoop Kumar[1], Navin Kamuni[1], and Madan Mohan Tito Ayyalasomayajula[1]

[1]Affiliation not available

April 18, 2024

Exploring Strategies for Privacy-Preserving Machine Learning in Distributed Environments

Suresh Dodda [1], Anoop Kumar [2], Navin Kamuni [3], Madan Mohan Tito Ayyalasomayajula[4]
[1]sureshr.dodda@gmail.com, [2]Anoop.kumar.2612@gmail.com, [3]navin.kamuni@gmail.com,
[4]mail2tito@gmail.com
[1] IT Department, Eudoxia Research center, USA
[2] IIT Roorkee, India
[3] AI-ML, BITS Pilani WILP, USA
[4] Aspen University, Phoenix, Arizona, USA

***Abstract*—Machine Learning (ML) with distributed privacy preservation is growing in significance as it focuses on facilitating multi-party learning without requiring actual data sharing. This is especially helpful for companies that want to work together but are unable to do so because of ethical, regulatory, or budgetary constraints on sharing data. In order to address these issues, this study examines three privacy-preserving algorithms: regularized logistic regression with Differential Privacy (DP), stochastic gradient descent (SGD) with differentially private updates, and a distributed Lasso that distributes gradients among data centers. The study emphasizes the relationship between error rate and privacy through these algorithms. In order to improve error rates for large datasets, both DP algorithms modify their sensitivity dependent on the amount of data, highlighting the significance of training data volume in model performance in the study. Results demonstrate that using the SGD; error rate can be reduced by employing random projections in advance.**

***Index Terms*—Distributed Privacy Preservation, Differential Privacy, Lasso, Machine Learning, Regularized Logistic Regression, Stochastic Gradient Descent**

I. Introduction

Big data, the Internet of Things (IoTs), and Machine Learning (ML) are at the forefront of technological advancements, leading to increased user awareness of their data trails and privacy concerns. Users are recognizing that their personal information, such as medical history or online activities, might be accessible to organizations or vulnerable to hackers, raising significant privacy issues. This concern is amplified by instances of organizations mishandling sensitive data, underscoring the need for secure data management practices. ML's application across various sectors, including healthcare for early disease detection and fraud detection in financial services, highlights its importance. However, the potential for collaborative innovation is often hampered by ethical, legal, and financial constraints on data sharing. For instance, sharing sensitive data between a pharmaceutical company and a healthcare organization might be unethical without ensuring privacy. This dilemma points to the necessity of integrating ML with privacy-preserving techniques, a challenge recognized by leading companies like Apple, Microsoft, and Google. Addressing this challenge, this study focuses on developing differential privacy techniques for distributed ML that allow for gradient sharing without direct data exchange, using the MNIST dataset to evaluate the proposed methods. The concept of anonymizing data, which involves removing identifiable information, has been deemed insufficient for protecting privacy due to the risk of re-identification from other available data sources. This was exemplified by the re-identification of William Weld's medical records from an anonymized dataset, despite the removal of direct identifiers. This incident underscores the limitations of traditional anonymization techniques and the need for more robust privacy-preserving methods.

This paper introduces differential privacy techniques for distributed ML, specifically distributed private Lasso and regularized logistic regression with Differential Privacy (DP), aiming to balance privacy and accuracy without requiring direct data sharing. The MNIST dataset serves as the testing ground for these algorithms, allowing for an in-depth analysis of the trade-offs between privacy protection and model performance. Structured as follows, the paper presents state-of-the-art work in Section II, details the study's materials and methods in Section III, analyzes the findings in Section IV, discusses the implications in Section V and concludes with conclusions and future research directions. This research contributes to the ongoing dialogue on privacy in the digital age, offering insights into the feasibility of ML applications that safeguard user privacy while maintaining accuracy.

II. State of the Art

Many different approaches have been attempted to preserve privacy. One of them is *k*-anonymity [12]. In *k* rows of the dataset, the goal is to make predefined identifiers homogeneous. To accomplish this, you can either change the identification to have a "range" or append a "*" when the identifiers differ. For instance, if age is used as an identifier, the file can indicate that the user was born in the years 1990–1999. An expansion of the *k*-anonymity approach is *l*-diversity [13]. The absence of variety in the sensitive attribute is why k-anonymity may leak information, leading to the development of

l-diversity, which requires at least l well-represented values for the sensitive attribute. T-closeness further enhances l-diversity by ensuring the distribution of the sensitive attribute in any equivalent class closely matches the distribution across the entire dataset. However, these techniques may not prevent security breaches in extreme cases, prompting the use of secure Multi-Party Computing (MPC) to preserve confidentiality. MPC allows multiple participants to compute a global function without revealing their input data. Alternatively, Federated Learning (FL) distributes a model across devices like smartphones and tablets, where it's trained locally. The results are then encrypted and shared with a cloud, which updates a global model without direct access to raw data, maintaining data privacy. Differential Privacy (DP) provides another layer of privacy by introducing noise and randomness to offer plausible deniability to users in the dataset. DP ensures that the dataset doesn't allow user identification, even if an adversary gains additional information, by utilizing randomization to protect users. DP supports population research while safeguarding user privacy, highlighting its role as a crucial concept in privacy preservation.

III. Materials and Methods

Convexity in ML optimization is crucial, ensuring that each local minimum is a global minimum, with strong convexity guaranteeing solution uniqueness and faster convergence rates, particularly vital in high-dimensional spaces for efficiently locating the global minimum. Privacy-preserving data analysis leverages DP, allowing insight extraction from datasets without compromising individual privacy. DP's inclusion of randomness, defined by parameters (ϵ, δ), minimizes the influence of any single data point on the analysis outcome, essential in sensitive data fields like healthcare and social science. Understanding a function's sensitivity is key to effective DP use, guiding the noise added to ensure privacy without significantly diminishing data utility. Depending on the desired privacy guarantees ((ϵ, δ)-DP) and function sensitivity, the choice between Laplace or Gaussian noise mechanisms is made, with Gaussian favored for assumed Gaussian data distributions and Laplace for its simplicity and strict ϵ-DP adherence. Regularization techniques like Lasso regression help prevent overfitting and assist in feature selection by penalizing large coefficients, crucial for models prone to interpreting training data noise as patterns. Dimensionality reduction methods such as PCA and random projection are indispensable in big data analytics, reducing variables to uncover underlying data patterns, enhancing model performance, and computational efficiency. Optimizing logistic regression with DP involves gradually adding noise, protecting training data individual privacy while maintaining predictive utility. SGD, combined with DP, offers privacy-preserving algorithms vital for sensitive application ML models, improved by gradient clipping to prevent large, potentially destabilizing steps.

Model evaluation strategies like hold-out and k-fold cross-validation ensure models are reliable predictors for new data, beyond training data performance. Hyperparameter optimization relies on grid search, systematically exploring a predefined grid to identify the hyperparameter combination that maximizes performance, a critical process for enhancing ML application effectiveness.

Data Analysis (MNIST)

A traditional dataset with a large number of well-known benchmarks is used. The dataset was created by [18]. It includes images of handwritten numbers together with their labels. One thousand test samples and sixty thousand training samples make up the dataset. Seventy-eight features are obtained from each 28 x 28 image. The target ranges from 0 to 9. A pixel in an image is represented by each feature. Pixel values range from 0 for black to 255 for white, with grey representing a mixture of black and white (e.g., grey = 127.5). Fig. 1 shows the images for each target. Since every pixel represents a different color, it's interesting to look at how those colors are distributed throughout all of the images. The majority of the data points are black, as can be seen when examining the sample figure (Fig. 1). Furthermore, since it only has spikes at the extremes, it is evident from that there is little to no grey in the images.

Fig. 1. An illustration of the images to forecast using every image has a label that is arranged numerically

After that, a study was conducted to find out how the images varied from the standard image. To calculate the usual image, the mean of each feature for each target was found, yielding 784 means—one for each dimension. The Euclidean distance between each image and the usual image was determined in order to examine the differences between each image and the typical one. The violin plot in Fig. 2 displays the results. The images in each category that deviate the most from the average are the ones with the highest scores. This indicates that the number that gets drawn the most frequently is 1. In contrast, the violin plot indicates that the categories with the highest mass, 0 and 2, appear to have the greatest diversity in their drawing. Finally, because image 8 has the largest spike, it seems to have the single worst drawing. The five images in each category with the largest Euclidean distance from the usual image are displayed in Fig. 3. Looking at the images makes it obvious that some of the numbers are quite badly drawn. Upon examining the images, it becomes evident that certain numbers are quite poorly painted. Moreover, a few of the numbers in the images, such the first 7 and 3, don't even resemble the actual numbers. To determine whether there was a class imbalance in the dataset, an easy calculation of the targets was performed (Table I demonstrates, there is imbalance).

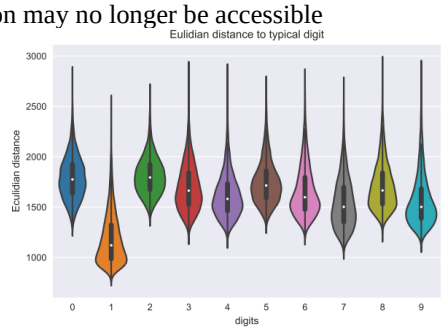


Fig. 2. Displays each image's Euclidean distance from its usual image in each category

Fig. 3. Displays the images that deviate the most from the average image

TABLE I
THE TOTAL NUMBER OF TARGET DATA POINTS FOR EVERY CATEGORY

Samples	0	1	2	3	4	5	6	7	8	9
Training	5923	6742	5958	6131	5842	5421	5918	6265	5851	5949
Test	980	1135	1032	1010	982	892	958	1028	974	1009

Data standardization is critical in ML as unstandardized data affects algorithms' performance, particularly in gradient-based methods like the distributed private Lasso. Standardizing gradients ensures step sizes are consistent, with the gradient indicating direction and the learning rate (η) determining step size. High-dimensionality (784 dimensions) complicates data visualization, necessitating PCA for dimension reduction. Despite standardization, PCA is essential to prevent larger-scale features from falsely appearing more significant. In Figure 4, projecting training samples onto the first two principal components reveals clustering of similar digits but only captures 9.7% of the original data's variance, highlighting the challenge in differentiating variables through PCA alone.

For binary classification problems, PCA visualizes the distinctions between digit pairs, with datasets for digits 4 and 9, and 0 and 1, analyzed separately. The PCA projection of 4 and 9 (Figure 5) shows overlapping classes, making class identification challenging without prior knowledge, and explains only 12.7% of variance. Conversely, the 0 and 1 projection demonstrates better class separation and accounts for 16.2% of the variance, suggesting easier differentiation between these digits. This contradicts the initial assumption that classifying between 4 and 9 is harder than 0 and 1, as the latter shows a more distinct distribution, particularly for digit 0, which has greater variation due to its visual characteristics in images. Figure 6 further supports this by showing how explained variance increases with more principal components for 0 and 1, indicating fewer components are needed to represent their data adequately.

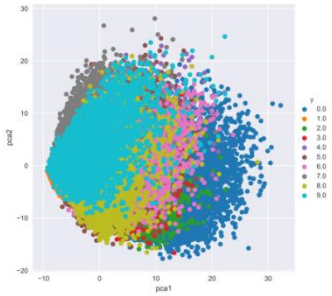


Fig. 4. PCA for every target variable

4 and 9 scattered onto their first two principal componenets.

0 and 1 scattered onto their first two principal componenets.

Fig. 5. A PCA of two datasets with only the labels 4 and 9 in one and only the labels

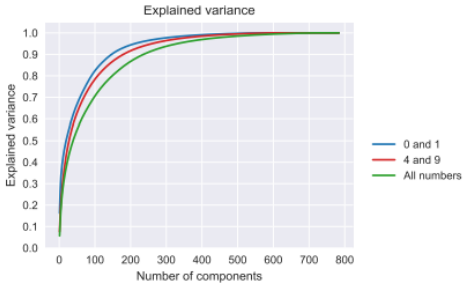


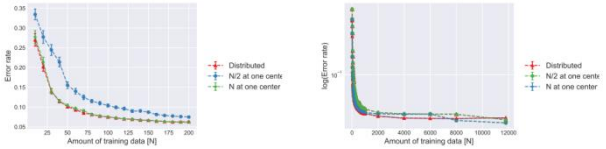
FIG. 6. DEMONSTRATES THE RISE IN THE EXPLAINED VARIANCE BY ADDING UP THE MAJOR COMPONENTS

IV. EXPERIMENTAL ANALYSIS

A. *Distributed Private Lasso*

This study develops and analyses a distributed private Lasso model, logistic regression with DP, and SGD with differentially private updates to examine the incorporation of privacy into ML. Showcasing novel approaches to privacy-preserving ML, the distributed private Lasso model is focused on binary classification issues using MNIST data (classifying digits 4 and 9 in particular). Two alternative scenarios were examined with this model: one in which half the data is analyzed by a single data center, and the other in which all the data is centralized. According to the hypothesis, when data was scarce, the distributed model would perform better than a single data center, but it might not be able to equal the performance of the completely centralized strategy. Extensive simulations utilizing grid search for hyperparameter tuning (weight decay λ and learning rate η) were performed to evaluate the models. Multiple iterations and 5-fold cross-validation were employed to ensure robustness. Error rates and feature selection skills were used to assess the models, and the distributed approach

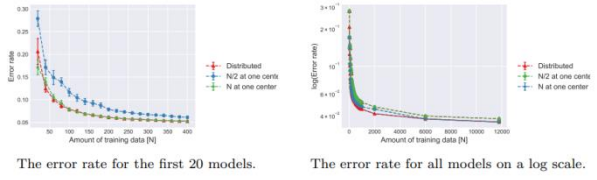
was anticipated to provide a balance between privacy and predictive performance. Two sets of tests were performed: one using PCA to reduce the data to the first 100 principal components and another utlizing the entire 784-dimensional data without PCA. The findings showed that although PCA sped up convergence and decreased processing requirements, it may make it more difficult for the model to choose features efficiently because PCA chooses features primarily on variance rather than predictive relevance. In contrast, the distributed Lasso model performed better in terms of error rate without PCA (refer to Fig. 7); in several cases, it even outperformed the centralized data model. This improvement was credited to the regularisation effect brought about by data distribution, which may lessen overfitting by placing some sort of restriction on the complexity of the model.



A close up of the first 20 data points. All results on log axis.

Fig. 7. Displays the PCA models' error rate

Furthermore, the distributed method without PCA made it possible to choose features with greater nuance, identifying both those that are most suggestive of the variations between the numbers 4 and 9 and those that have little to no variance, which are judged useless for classification. In applications where interpretability is just as critical as accuracy, this fine-grained feature selection is essential for comprehending the model's decision-making process. The possibility of reverse engineering the aggregated gradient to extract details about specific data points was investigated in order to allay privacy concerns (refer to Figs. 8, 9, and 10). This experiment highlights the benefits and drawbacks of distributed methods, adding to the larger subject of privacy-preserving ML. In order to reduce the hazards associated with reverse engineering, further study could improve privacy safeguards, investigate alternate dimensionality reduction strategies, and further refine these models.



The error rate for the first 20 models. The error rate for all models on a log scale.

Fig. 8. Displays the without PCA models' error rate

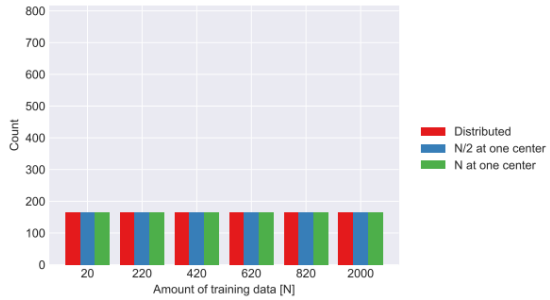


Fig. 9. The quantity of weights whose absolute value is lower

shows a image of 4 shows a image of 9

Fig. 10. Displays the distributed Lasso feature selection with the most data

B. Regularized Logistics Regression with Differential Privacy

The difficult classification between digits 4 and 9 from the MNIST dataset is the study's main emphasis as it examines a regularized binary logistic regression model with DP. To accomplish DP, the model's weights are supplemented with Laplacian noise, the amount of which is dictated by the function's sensitivity and the privacy parameter ϵ. The sensitivity decreases as N or λ grows, indicating that larger datasets or more regularization contribute to privacy by lessening the impact of individual data points. The sensitivity is inversely proportional to the total number of observations (N) and directly to the weight decay (λ). Therefore, logistic regression models with l_2 norm regularization were trained on a range of data sizes, from $1/50^{th}$ to the entire dataset. Next searches were conducted based on discovered ranges to fine-tune the initial broad grid searches for the ideal λ. To verify randomness and determine the most widely used ideal parameters, the models were tested throughout a number of simulations. The outcomes of the trial demonstrate the balance between accuracy and privacy. The model's error rate falls as ϵ rises, signifying less noise, which is consistent with the hypothesis that less noise reduces model accuracy (refer to Fig. 11). On the other hand, the error rates that were similar to random guessing were created by the lowest values of ϵ, which emphasizes how much noise impacts the performance of the model. However, the influence of the additional noise decreases with increasing training data, improving model accuracy (refer to Fig. 12). This is explained by the fact that as N increases, sensitivity decreases and less noise can be added.

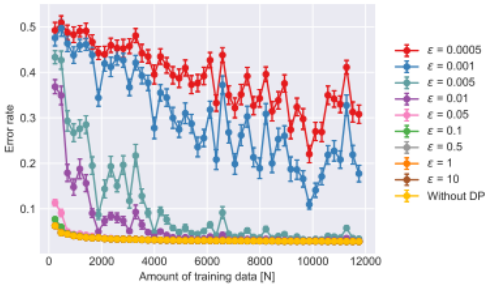
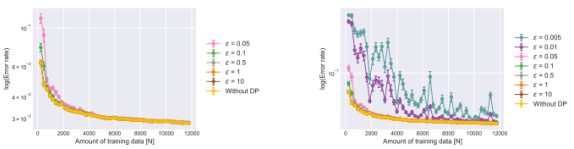


Fig. 11. Displays the logistic regression error rate with and without DP



The best 6 accuracy's. The best 8 accuracy's.

Fig. 12. An enlargement of the error rates on a log axis in Fig. 11

The study also showed that models with ϵ values from 0.5 and higher converged to error rates that were similar to non-private logistic regression models across a significant amount of the data spectrum (refer to Fig. 13). This implies that the accuracy of the model is not significantly impacted by a moderate level of noise. But it becomes important to distinguish between the privacy levels provided by various ϵ values since even minor variations in ϵ might result in considerable variations in privacy assurances. The study emphasizes how crucial it is to choose ϵ carefully, striking a balance between model accuracy and privacy needs. Higher ϵ values decrease the privacy guarantee but may provide better accuracy by adding less noise. The results indicate that it is possible to obtain strong privacy safeguards (lower ϵ) without sacrificing model performance if enough data is available. In a nutshell, our study sheds insight on the complex interplay among data privacy, model precision, and training data volume within the framework of ML-DP. It validates the feasibility of upholding privacy restrictions without sacrificing model efficacy, particularly when dealing with larger datasets. It is still crucial to choose ϵ carefully, weighing the demands of privacy against accuracy objectives. This decision is context-specific, changing depending on the particular specifications and application-specific sensitivity.

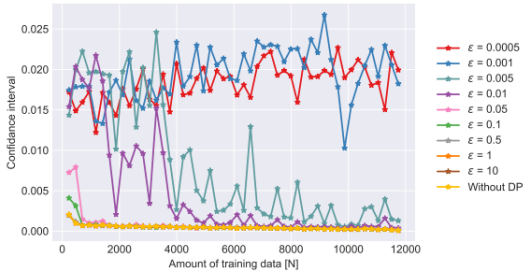


Fig. 13. Displays the error rate's confidence interval

C. *Stochastic Gradient Descent with Differential Privacy*

This study presents a nuanced approach to privacy preservation and model accuracy optimization through the investigation of SGD with DP for logistic regression on the MNIST dataset, specifically for the classification of digits 4 and 9. DP allows the SGD algorithm to adjust to protect specific data points during training while minimizing the model's predictive power. Laplacian noise is added to the weights to mask individual contributions. Normalizing predictors, mapping targets to -1 and 1, projecting data onto a unit ball, and randomly projecting dimensions to 50 dimensions were all part of the preprocessing steps. The process of reducing dimensionality not only speeds up computations but also naturally caps the overall noise introduced in each gradient update, balancing privacy protection with sufficient data accuracy to enable precise categorization. Hyperparameter tweaking is required under privacy constraints, as proven by the training of fifty SGD models with different data sizes. Gradient clipping is used to prevent the gradient's norm from exceeding a predefined threshold, preserving the privacy guarantee. The DP model necessitates careful consideration of the learning rate, batch size, and weight decay (refer to Fig. 14). Grid search was used to guide the hyperparameter selection process, with a focus on bigger batch sizes to minimize noise-induced variance and closer resemble the genuine gradient.

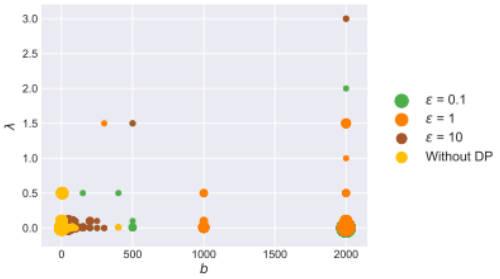


Fig. 14. A scatter plot with the λ and η separated

The outcomes of the experiment demonstrated the trade-offs and inherent difficulties between model accuracy and privacy levels (measured by ϵ). As the quantity of training data rose, models trained with higher ρ values—which indicate less noise—became closer to the accuracy of non-private SGD models. On the other hand, higher error rates were associated with lower ϵ values, which indicated the effect of increased noise on model performance (refer to Figs. 15 and 16). This shows how privacy protection and model accuracy are directly correlated, with lower ϵ values corresponding to greater privacy and thus fewer accurate predictions. The intricate relationship between the quantity of features, the degree of DP used, and the final model accuracy was further highlighted by dimensionality analysis. Models with lower dimensions and suitable ϵ values showed that privacy and model efficacy could be maintained through a precisely calibrated feature space reduction.

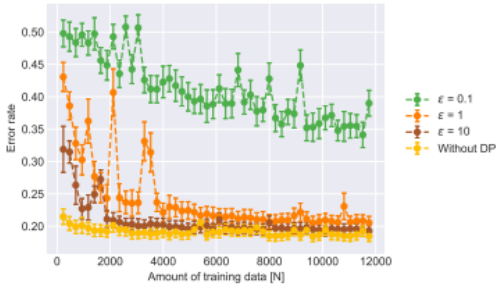


Fig. 15. The SGD's error rate when using DP

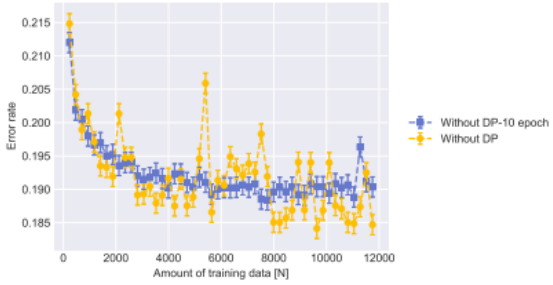


Fig. 16. The SGD without DP error rate displayed with varying epoch counts

V. Conclusion and Future Works

Using three independent approaches—distributed Lasso, logistic regression with DP, and SGD with differentially private updates—this study investigates the relationship between privacy and ML. Every approach presents a different angle on how to balance privacy protection and ML needs, emphasizing the connection between privacy, error rates, and data volume.

The distributed Lasso method highlights the benefits of distributing gradients across data centers, particularly when data is scarce, showcasing the advantage of collaborative environments for enhancing model accuracy and convergence. Differentially private logistic regression illustrates how increasing data volume can maintain privacy without sacrificing accuracy, managing the trade-off between privacy and performance by adjusting the privacy budget (ϵ). SGD with differentially private updates introduces local differential privacy, focusing on batch size, epoch count, and the dimensionality-privacy link. Despite the highest error rates, this method provides insights into integrating privacy with iterative learning algorithms, underscoring the growing importance of privacy-preserving ML as data collection expands. In parallel to our exploration of privacy-preserving strategies in distributed ML environments, there have been significant strides in enhancing end-to-end multi-task dialogue systems [19], offering a comprehensive approach to improving interaction capabilities while potentially navigating privacy concerns inherent in user data handling. Similarly, advancing audio fingerprinting accuracy [20], particularly in mitigating challenges posed by background noise and distortion, underscores the necessity of sophisticated data processing techniques that can be aligned with privacy-preserving mechanisms. Future research could explore the algorithms' performance on multiclass tasks and different datasets, assess the integration of DP into the distributed Lasso for potentially better generalization, and consider the logistic regression model with DP in a distributed framework to examine privacy implications.

VI. Declarations

A. ***Funding:*** No funds, grants, or other support was received.

B. ***Conflict of Interest:*** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

C. ***Data Availability:*** Data will be made on reasonable request.

D. Code Availability: Code will be made on reasonable request.

References

[1] M. Kanojia, P. Kamani, G. S. Kashyap, S. Naz, S. Wazir, and A. Chauhan, "Alternative Agriculture Land-Use Transformation Pathways by Partial-Equilibrium Agricultural Sector Model: A Mathematical Approach," Aug. 2023, Accessed: Sep. 16, 2023. [Online]. Available: https://arxiv.org/abs/2308.11632v1

[2] G. S. Kashyap, A. E. I. Brownlee, O. C. Phukan, K. Malik, and S. Wazir, "Roulette-Wheel Selection-Based PSO Algorithm for Solving the Vehicle Routing Problem with Time Windows," Jun. 2023, Accessed: Jul. 04, 2023. [Online]. Available: https://arxiv.org/abs/2306.02308v1

[3] G. S. Kashyap, K. Malik, S. Wazir, and R. Khan, "Using Machine Learning to Quantify the Multimedia Risk Due to Fuzzing," *Multimedia Tools and Applications*, vol. 81, no. 25, pp. 36685–36698, Oct. 2022, doi: 10.1007/s11042-021-11558-9.

[4] G. S. Kashyap, A. Siddiqui, R. Siddiqui, K. Malik, S. Wazir, and A. E. I. Brownlee, "Prediction of Suicidal Risk Using Machine Learning Models." Dec. 25, 2021. Accessed: Feb. 04, 2024. [Online]. Available: https://papers.ssrn.com/abstract=4709789

[5] G. S. Kashyap, D. Mahajan, O. C. Phukan, A. Kumar, A. E. I. Brownlee, and J. Gao, "From Simulations to Reality: Enhancing Multi-Robot Exploration for Urban Search and Rescue," Nov. 2023, Accessed: Dec. 03, 2023. [Online]. Available: https://arxiv.org/abs/2311.16958v1

[6] G. S. Kashyap *et al.*, "Detection of a facemask in real-time using deep learning methods: Prevention of Covid 19," Jan. 2024, Accessed: Feb. 04, 2024. [Online]. Available: https://arxiv.org/abs/2401.15675v1

[7] H. Habib, G. S. Kashyap, N. Tabassum, and T. Nafis, "Stock Price Prediction Using Artificial Intelligence Based on LSTM– Deep Learning Model," in *Artificial Intelligence & Blockchain in Cyber Physical Systems: Technologies & Applications*, CRC Press, 2023, pp. 93–99. doi: 10.1201/9781003190301-6.

[8] S. Wazir, G. S. Kashyap, and P. Saxena, "MLOps: A Review," Aug. 2023, Accessed: Sep. 16, 2023. [Online]. Available: https://arxiv.org/abs/2308.10908v1

[9] N. Marwah, V. K. Singh, G. S. Kashyap, and S. Wazir, "An analysis of the robustness of UAV agriculture field coverage using multi-agent reinforcement learning," *International Journal of Information Technology (Singapore)*, vol. 15, no. 4, pp. 2317–2327, May 2023, doi: 10.1007/s41870-023-01264-0.

[10] S. Wazir, G. S. Kashyap, K. Malik, and A. E. I. Brownlee, "Predicting the Infection Level of COVID-19 Virus Using Normal Distribution-Based Approximation Model and PSO," Springer, Cham, 2023, pp. 75–91. doi: 10.1007/978-3-031-33183-1_5.

[11] A. Bruckman, "Studying the amateur artist: A perspective on disguising data collected in human subjects research on the Internet," *Ethics and Information Technology*, vol. 4, no. 3, pp. 217–231, 2002, doi: 10.1023/A:1021316409277.

[12] L. Sweeney, "Achieving k-anonymity privacy protection using generalization and suppression," *International Journal of Uncertainty, Fuzziness and Knowldege-Based Systems*, vol. 10, no. 5, pp. 571–588, May 2002, doi: 10.1142/S021848850200165X.

[13] A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam, "ℓ-diversity: Privacy beyond k-anonymity," *ACM Transactions on Knowledge Discovery from Data*, vol. 1, no. 1, Mar. 2007, doi: 10.1145/1217299.1217302.

[14] L. Ninghui, L. Tiancheng, and S. Venkatasubramanian, "t-Closeness: Privacy beyond k-anonymity and ℓ-diversity," in *Proceedings - International Conference on Data Engineering*, 2007, pp. 106–115. doi: 10.1109/ICDE.2007.367856.

[15] Y. Lindell, "Secure Multiparty Computation for Privacy Preserving Data Mining," in *Encyclopedia of Data Warehousing and Mining*, IGI Global, 2011, pp. 1005–1009. doi: 10.4018/978-1-59140-557-3.ch189.

[16] A. Ghosh, J. Chung, D. Yin, and K. Ramchandran, "An efficient framework for clustered federated learning," in *Advances in Neural Information Processing Systems*, 2020, vol. 2020-Decem. Accessed: Mar. 17, 2022. [Online]. Available: https://proceedings.neurips.cc/paper/2020/hash/e32cc80bf07915058ce90722ee17bb71-Abstract.html

[17] S. P. Kasiviswanathan, O. K. Lee, K. Nissim, S. Raskhodnikova, and A. Smith, "What can we learn privately?," in *SIAM Journal on Computing*, Jun. 2011, vol. 40, no. 3, pp. 793–826. doi: 10.1137/090756090.

[18] Y. LeCun, C. Cortes, and C. J. C. Burges, "The MNIST database of handwritten digits, 1998," *URL http://yann. lecun. com/exdb/mnist*, vol. 10, no. 34, p. 14, 1998, Accessed: May 20, 2023. [Online]. Available: https://cir.nii.ac.jp/crid/1571417126193283840

[19] N. Kamuni, H. Shah, S. Chintala, N. Kunchakuri and S. Alla, "Enhancing End-to-End Multi-Task Dialogue Systems: A Study on Intrinsic Motivation Reinforcement Learning Algorithms for Improved Training and Adaptability," 2024 IEEE 18th International Conference on Semantic Computing (ICSC), Laguna Hills, CA, USA, 2024, pp. 335-340, doi: 10.1109/ICSC59802.2024.00063

[20] N. Kamuni, S. Chintala, N. Kunchakuri, J. Narasimharaju and V. Kumar, "Advancing Audio Fingerprinting Accuracy with AI and ML: Addressing Background Noise and Distortion Challenges," in 2024 IEEE 18th International Conference on Semantic Computing (ICSC), Laguna Hills, CA, USA, 2024 pp. 341-345. doi: 10.1109/ICSC59802.2024.00064