FEDERATED QUANTUM MACHINE LEARNING WITH DIFFERENTIAL PRIVACY

Rod Rofougaran[1], *Shinjae Yoo*[2], *Huan-Hsin Tseng*[2], *Samuel Yen-Chi Chen*[3]

[1]School of Applied and Engineering Physics, Cornell University, Ithaca, NY 14853
[2]Computational Science Initiative, Brookhaven National Laboratory, Upton, NY 11793
[3]Wells Fargo, New York, NY 10017

ABSTRACT

Privacy preservation is essential in deploying artificial intelligence with sensitive data. Quantum computations offer enhanced security, leveraging the no-cloning theorem to make them an ideal choice for secure computing. Prior research has explored Quantum Federated Learning (QFL) and Quantum Differential Privacy (QDP) separately, but combining these has not yet been studied. Our work integrates these techniques on a quantum platform to provide robust protection against data breaches and model inversion attacks, significantly boosting AI security and efficiency. We demonstrate this by successfully classifying the Cats vs Dogs dataset on a quantum-classical hybrid model, achieving over 98% accuracy and maintaining epsilon values under 1.3. Our results validate the effectiveness of federated, differentially private training on Noisy Intermediate-Scale Quantum (NISQ) devices.

1. MOTIVATION

Quantum computing demonstrates potential advantages over classical systems in specialized, complex tasks essential to the evolution of quantum technology. Their security, reinforced by the no-cloning principle, ensures robust protection against unauthorized data access. Since there are limitations in current NISQ devices, we use the variational quantum algorithms (VQA) to facilitate computations in limited qubits.

This work aims to achieve quantum model security via the learning process. We join the merits of two distinct privacy-preserving classical techniques: Federated Learning (FL) and Differential Privacy (DP). As a result, we can effectively shield against both model inversion attacks and data leakage, while operating on an inherently secure quantum platform.

2. APPROACH

There are three major components in this work to be integrated.

Federated Learning (FL)

FL [1] processes extensive datasets by distributing tasks across multiple nodes, decentralizing training data among clients. It begins with initializing a global model $\Theta \in \mathbb{R}^n$ and distributing identical copies $\Theta_1, \ldots, \Theta_K$ to K clients, where $\Theta = \Theta_1 = \cdots = \Theta_K$. Each client j from $[K] = \{1, \ldots, K\}$ independently trains its model Θ_j to produce updated models $\widetilde{\Theta}_j \neq \Theta_j$. These are aggregated into a new global model $\widetilde{\Theta}$, cycling through several iterations. This decentralized approach enhances security and suits quantum machine learning on NISQ devices, which efficiently process smaller datasets. Notably, quantum federated learning without differential privacy maintains testing accuracy [2], illustrated in Fig. 1.

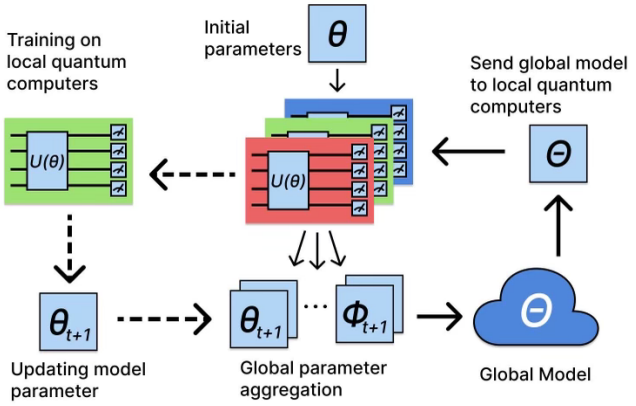

Fig. 1. The concept of QFL.

Differential Privacy (DP)

Considering two datasets: one containing X and the other without X, it is crucial that the difference in their model outputs is limited to a specific bound, ϵ. Without such a restriction, an individual with access to the model could potentially deduce whether X was included in the dataset.

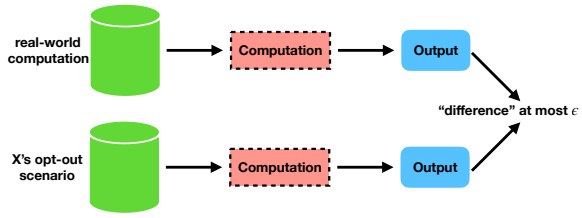

Fig. 2. The concept of differential privacy.

The ϵ-differential private definition follows from [3]:

Definition 2.1. *Let $\mathcal{M}$ be a randomized algorithm whose (functional) image is a collection of (probabilistic)* events *$\mathcal{S}$ and the domain is a collection of datasets. If $\mathcal{M}$ is said to be (ϵ, δ)-differentially private for any dataset $\mathcal{D}_1$, $\mathcal{D}_2$ that differ*

on a single data point (denoted as $||\mathcal{D}_1| - |\mathcal{D}_2|| = 1$), we have

$$Pr[\mathcal{M}(\mathcal{D}_1) \in \mathcal{S}] \leq e^{\epsilon} \cdot Pr[\mathcal{M}(\mathcal{D}_2) \in \mathcal{S}] + \delta \qquad (1)$$

The quantity $\delta \geq 0$ has the meaning of *failure probability* [3], while ϵ indicates an upper bound on the privacy loss. By adding Gaussian noise and gradient clipping into the training optimization [4], a classifier can be made to guarantee differential privacy Eq. (1). In [4], a fact is proved:

Theorem 2.1. *There exists c_1 and c_2 so that given the number of epochs T and the sampling probability $q = L/N$ where L is the batch size and N is the total number of examples, for any $\epsilon < c_1 q^2 T$, randomized algorithm $\mathcal{M}$ is (ϵ, δ)-differentially private for any $\delta > 0$ if we choose the noise level σ:*

$$\sigma \geq \frac{c_2 q \sqrt{T \log\left(\frac{1}{\delta}\right)}}{\epsilon}$$

The value of ϵ is a function of the following training parameters: the total number of examples, batch size, noise multiplier, number of epochs, and δ. The main correlation is the inverse relationship of ϵ and the noise injected to the input.

Variational Quantum Circuits (VQC)

Variational quantum circuits (VQC) serve as the quantum counterpart to the classical neural networks. A VQC takes three major steps to learn data: (1) the *encoding* part, translating a classical vector $\mathbf{x} \in \mathbb{R}^m$ into a quantum state $|\xi\rangle$ by an embedding function $\mathbf{x} \mapsto E(\mathbf{x})$ so that $|\xi\rangle = E(\mathbf{x}) |0\rangle^{\otimes n}$ (see Fig. 3). In this work, we follow the procedure in [2]. (2) a *learnable* (*variational*) quantum gate $W(\phi)$ including multiple single-qubit rotations such that $W_{ij}(\phi^{(ij)}) = e^{i(\sigma_x \alpha_{ij} + \sigma_y \beta_{ij} + \sigma_z \gamma_{ij})}$. Here, σ_k's are Pauli matrices with i, j as the variational block and qubits index and $\phi_{ij} = (\alpha_{ij}, \beta_{ij}, \gamma_{ij}) \in \mathbb{R}^3$ as the corresponding learnable parameters. The final step: (3) measurement operations to retrieve the circuit information, where the Pauli-Z is utilized for the expectation values in this work. Collectively, the three steps give us a learnable quantum function $\overrightarrow{f(\mathbf{x}; \phi)} = \left(\left\langle \hat{Z}_1 \right\rangle, \cdots, \left\langle \hat{Z}_N \right\rangle\right)$ with $\left\langle \hat{Z}_k \right\rangle = \left\langle 0 \left| E^{\dagger}(\mathbf{x}) W^{\dagger}(\phi) \hat{Z}_k W(\phi) E(\mathbf{x}) \right| 0 \right\rangle$. By *varying* parameters ϕ, the minimization of the objective function can be achieved at $\phi^* = \text{argmin}_{\phi} \mathcal{L}(f(\mathbf{x}; \phi))$ where $\mathcal{L}$ is the loss function.

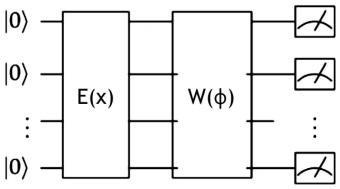

Fig. 3. **A generic structure of a VQC.** A VQC comprises an encoding module denoted as $E(\mathbf{x})$, a trainable component represented as $W(\phi)$, and subsequent measurement operations.

The VQC circuit is known to enhance differential privacy when encoding classical datasets [5] and exhibit greater expressiveness than classical neural networks [6, 7, 8, 9]. VQCs can also be trained on smaller datasets with high efficiency [10] and are applied across various domains like classification, reinforcement learning, natural language processing, and sequence modeling.

We integrate the DP and FL in QML via VQC to form our proposed method as **Algorithm 1**.

Algorithm 1 QFL-DP

Input: Examples $\{x_1, \ldots, x_M\}$, loss function $\mathcal{L}(\theta) = \frac{1}{N} \sum_i \mathcal{L}(\theta, x_i)$.

Parameters: Clients K, selected J, local epochs T, rounds R, learning rate η_t, noise scale σ, group size L, gradient norm bound C.

Partition: From M examples, construct $\mathcal{D}_1, \ldots, \mathcal{D}_K$ among K clients randomly, $|\mathcal{D}_i| = N = M/K$

Initialize: Quantum global model $\Theta_0 \in \mathbb{R}^n$

for $r \in [R]$ **do**
  Model distribution:
  Make K identical copies of Θ_r for local set
  $\{\Phi_{r1}, \ldots, \Phi_{rK}\}$ and send Φ_{rk} to client k
  Take random sample J from K clients
  for $j \in [J]$ **do**
    for $t \in [T]$ **do**
      DP client update:
      Perform DP-SGD$(N, \mathcal{L}, \eta_t, \sigma, L, C)$ on
      $\Phi_{rj} \leftarrow \widetilde{\Phi}_{rj} \neq \Phi_{rj}$
    end for
  end for
  Model aggregation:
  $\Theta_{r+1} =$ averaging the parameters across
  each model in $\{\widetilde{\Phi}_{rj}\}_{j=1}^{J}$
end for

Output: Θ_R and compute the overall privacy cost (ϵ, δ) using a privacy accounting method.

It is worth mentioning, due to the current constraints in NISQ devices we incorporate classical networks (pre-trained VGG16 model) for dimension reduction, prior to feeding data into a VQC so that our utilization of VQC becomes a *hybrid quantum-classical transfer learning* [11], see Fig. 4.

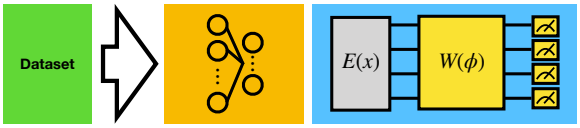

Fig. 4. **Hybrid Quantum-Classical transfer learning.**

3. EXPERIMENTS, RESULTS AND IMPACT

Experiments

We use 23,000 Cats vs Dog images [12] to demonstrate the proposed QFL across 100 clients. Training occurs in rounds, with randomly selected groups of 5 clients. At the start of each round, the global model is shared with all clients, but

only the chosen 5 perform local SGD training for a set number of epochs. The parameters from these selected clients are aggregated to update the global model for the next round. We validate our the QFL by exploring different settings, including varying the number of local epochs (1, 2, and 4) and incorporating a non-differentially private model. Each training process is repeated three times to average the outputs and reduce variance. Additionally, we conduct experiments to assess the impact of noise levels during training.

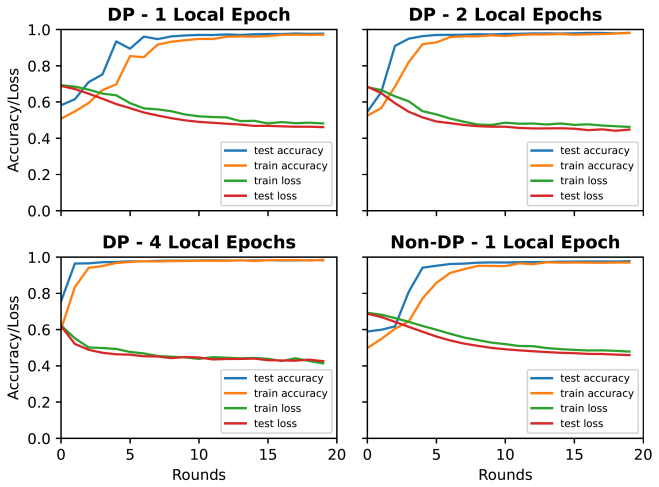

Fig. 5. All DP plots are $(\epsilon = 1.24, \delta = 10^{-5})$-DP and acquire test accuracy converging at approximately 0.98.

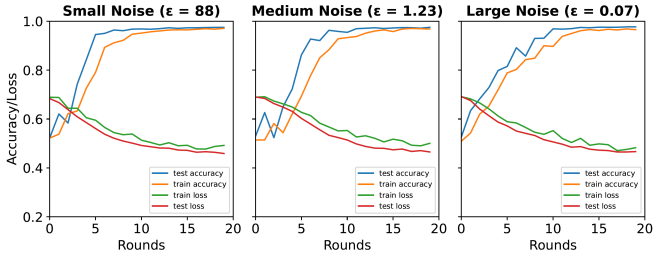

Fig. 6. [From left to right, $\sigma = 0.15, 1, 4$] All plots indicate test accuracy convergence at approximately 0.98.

Results

QFL-DP with different local epochs We first compare the results of QFL with DP training with various local epochs and the non-DP QFL. The results are shown in Fig. 5. We observe that all of our models converge to test accuracies of approximately 0.98 with ϵ's hovering around 1.24. It is important to note that the epsilon calculated was the global one, which is a function of total rounds. We also observe that as local epochs increase, a reduction in the number of rounds required to reach convergence, with a decline in variance. Finally, we observe that differentially private training converges slower and with higher variance, which aligns with expectations attributed to the introduction of noise. Additionally, our results are consistent with those of Chen et al. [2], which show that the testing accuracy and loss of federated training approximately converge to that of non-federated training.

QFL-DP with different noise levels We further study the correlation between the loss of privacy bound and the accuracy/loss of our models. We study the impact of noise via the increase in σ or equivalently the decrease in ϵ. As shown in Fig. 6, higher ϵ results in a slower, higher-variance training process. Generally, increasing the noise enhances privacy but will decrease classification accuracy. However, our results show that the final accuracies of the three cases are not different. Possible reasons are the simplicity of our Cats vs Dogs example and the capabilities of our model architecture.

Impact

Our work demonstrates the effectiveness of differentially private quantum federated learning in mitigating privacy concerns while maintaining competitive performance for NISQ devices. We recognize the need to explore more complex tasks tailored to quantum algorithms and conduct comparative assessments against classical methods to advance the field of privacy-preserving quantum machine learning.

4. REFERENCES

[1] Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Artificial intelligence and statistics*. PMLR, 2017, pp. 1273–1282.

[2] Samuel Yen-Chi Chen and Shinjae Yoo, "Federated quantum machine learning," *Entropy*, vol. 23, no. 4, pp. 460, 2021.

[3] Cynthia Dwork, Aaron Roth, et al., "The algorithmic foundations of differential privacy," *Foundations and Trends® in Theoretical Computer Science*, vol. 9, no. 3–4, pp. 211–407, 2014.

[4] Martin Abadi, Andy Chu, Ian Goodfellow, H Brendan McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang, "Deep learning with differential privacy," in *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security*, 2016, pp. 308–318.

[5] Armando Angrisani, Mina Doosti, and Elham Kashefi, "Differential privacy amplification in quantum and quantum-inspired algorithms," *arXiv preprint arXiv:2203.03604*, 2022.

[6] Sukin Sim, Peter D Johnson, and Alán Aspuru-Guzik, "Expressibility and entangling capability of parameterized quantum circuits for hybrid quantum-classical algorithms," *Advanced Quantum Technologies*, vol. 2, no. 12, pp. 1900070, 2019.

[7] Trevor Lanting, Anthony J Przybysz, A Yu Smirnov, Federico M Spedalieri, Mohammad H Amin, Andrew J Berkley, Richard Harris, Fabio Altomare, Sergio Boixo, Paul Bunyk, et al., "Entanglement in a quantum annealing processor," *Physical Review X*, vol. 4, no. 2, pp. 021041, 2014.

[8] Yuxuan Du, Min-Hsiu Hsieh, Tongliang Liu, and Dacheng Tao, "The expressive power of parameterized quantum circuits," *arXiv preprint arXiv:1810.11922*, 2018.

[9] Amira Abbas, David Sutter, Christa Zoufal, Aurélien Lucchi, Alessio Figalli, and Stefan Woerner, "The power of quantum neural networks," *Nature Computational Science*, vol. 1, no. 6, pp. 403–409, 2021.

[10] Matthias C Caro, Hsin-Yuan Huang, Marco Cerezo, Kunal Sharma, Andrew Sornborger, Lukasz Cincio, and Patrick J Coles, "Generalization in quantum machine learning from few training data," *Nature communications*, vol. 13, no. 1, pp. 1–11, 2022.

[11] Andrea Mari, Thomas R Bromley, Josh Izaac, Maria Schuld, and Nathan Killoran, "Transfer learning in hybrid classical-quantum neural networks," *Quantum*, vol. 4, pp. 340, 2020.

[12] Jeremy Elson, John (JD) Douceur, Jon Howell, and Jared Saul, "Asirra: A captcha that exploits interest-aligned manual image categorization," in *Proceedings of 14th ACM Conference on Computer and Communications Security (CCS)*. October 2007, Association for Computing Machinery, Inc.