

Quantum Private Distributed Learning Through Blind Quantum Computing

Weikang Li,^{1,*} Sirui Lu,^{1,2,*} and Dong-Ling Deng^{1,3,†}

¹Center for Quantum Information, IIS, Tsinghua University, Beijing 100084, People's Republic of China

²Max-Planck-Institut für Quantenoptik, Hans-Kopfermann-Str. 1, D-85748 Garching, Germany

³Shanghai Qi Zhi Institute, 41th Floor, AI Tower, No. 701 Yunjin Road, Xuhui District, Shanghai 200232, China

Private distributed learning studies the problem of how multiple distributed entities collaboratively train a shared deep network with their private data unrevealed. With the security provided by the protocols of blind quantum computation, the cooperation between quantum physics and machine learning may lead to unparalleled prospect for solving private distributed learning tasks. In this paper, we introduce a quantum protocol for distributed learning that is able to utilize the computational power of the remote quantum servers while keeping the private data safe. For concreteness, we first introduce a protocol for private single-party delegated training of variational quantum classifiers based on blind quantum computing and then extend this protocol to multiparty private distributed learning incorporated with differential privacy. We carry out extensive numerical simulations with different real-life datasets and encoding strategies to benchmark the effectiveness of our protocol. We find that our protocol is robust to experimental imperfections and is secure under the gradient attack after the incorporation of differential privacy. Our results show the potential for handling computationally expensive distributed learning tasks with privacy guarantees, thus providing a valuable guide for exploring quantum advantages from the security perspective in the field of machine learning with real-life applications.

Quantum machine learning is an emergent interdisciplinary field that explores the interplay between machine learning and quantum physics [1–3]. Recently, it has attracted tremendous attentions across different communities. On the one hand, machine learning, or more broadly artificial intelligence [4], has achieved momentous success over the past decade [5, 6] and a number of notoriously challenging problems, such as playing the game of Go [7, 8] or predicting protein structures [9], have been cracked successfully. This brings new intriguing opportunities for utilizing machine learning to tackle outstanding problems in quantum science [2, 3, 10]. On the other hand, as manifested by the experimental demonstration of quantum supremacy [11, 12], the field of quantum computing [13] has also made remarkable progress in recent years, giving rise to new striking possibilities of enhancing machine learning with quantum devices, in turn. Along this direction, a variety of works have shown strong evidences that quantum computers could outperform classical computers in solving certain machine learning problems [1–3]. Notable examples include the Harrow-Hassidim-Lloyd algorithm [14], quantum principal component analysis [15], quantum-enhanced feature space [16, 17], quantum generative models [18–20], and quantum support vector machines [21], etc. Most of these existed works study the possible quantum advantages from the perspective of quantum speedup in machine learning. In this paper, we explore possible quantum advantages from the security perspective and introduce a quantum protocol for private distributed learning based on the idea of blind quantum computing [22–30] (see Fig. 1 for a pictorial illustration).

Private distributed learning studies the interplay between privacy and machine learning, which is of crucial importance nowadays in the data-intensive and sensitive society [31]. Indeed, in many collaborative learning tasks [32, 33], it is highly desirable that the privacy for each party is preserved. To this end, a popular proposal developed in recent years concerns federated learning [34, 35], where given a share ma-

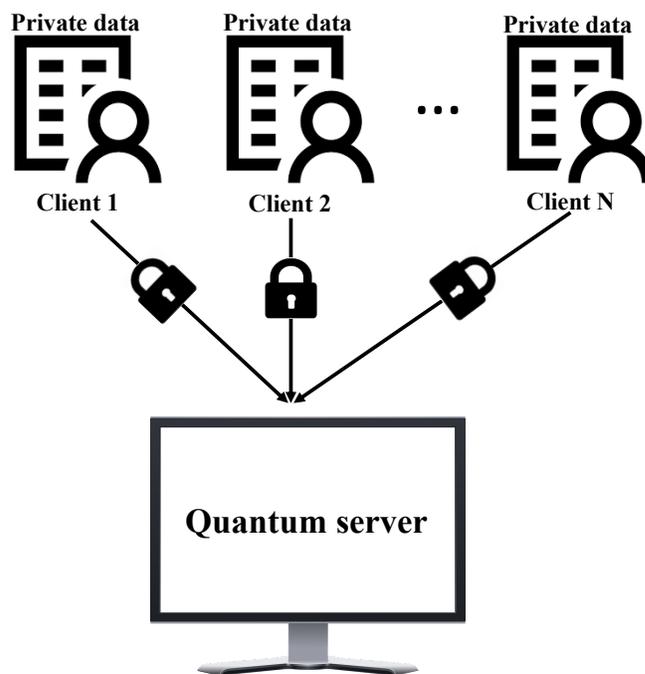


FIG. 1. Schematic illustration of private distributed learning with a quantum server, where multiple clients can utilize the computational power from the quantum server to collaboratively train a shared model while leaking no information about their private data during the computing process.

chine learning model, the updates of the model rather than the sensitive data are uploaded to the public server. In addition, to better fight against the malicious attacks that try to recover the sensitive data from the updated parameters of the model (such as the gradients), the differential privacy method that adds noise to the shared updates have been extensively studied over the years [36–38]. More recently, another approach to private distributed learning has also been introduced (dubbed

InstaHide), which encrypts each training image with a "one-time secret key" consisting of mixing several randomly chosen images and applying a random pixel-wise mask [39]. However, each of these approaches bears its pros and cons, and exhibits different levels of security. For instance, the differential privacy method often comes at the cost of accuracy and does not apply to side-channel computations performed by malicious parties [40, 41]. Fully-homomorphic encryption [42] can guarantee privacy against arbitrary side-computations by adversary during training. Yet, it is impractical in modern deep learning settings due to the high computational overheads and requirement of special setups.

In this paper, we introduce a protocol for private distributed learning based on blind quantum computing [22–30]. We first introduce a protocol for private single-party delegated training with variational quantum classifiers, and then extend this protocol to multiparty distributed learning as shown in Fig. 1, where there are several parties that wish to collaboratively train a shared model, e.g., several hospitals with inadequate and sensitive patient data on each side want to train a shared model for disease diagnosis. To better protect the privacy for each client, we incorporate the differential privacy method into the protocol. We carry out extensive numerical simulations with different real-life datasets (such as the MNIST dataset about handwritten digit images and the Wisconsin Diagnostic Breast Cancer dataset) and encoding strategies to benchmark the effectiveness of our protocol. In addition, we also demonstrate that our protocol is robust to experimental imperfections and is secure under the renowned gradient attack after the incorporation of differential privacy. We stress two intriguing advantages of our quantum private distributed learning protocol compared with its classical counterparts, such as federated learning. First, in our protocol the clients can delegate the computation of the gradient to the server without trusting the server. Second, we may achieve exponential speedup for some specific learning tasks. Our results open a new avenue for exploring quantum advantages in tackling practical machine learning tasks, which would be crucial for achieving large-scale unconditionally secure private distributed learning with future quantum technologies.

The general recipe.—To start with, we first recap the protocol of universal blind quantum computation (UBQC) introduced in [23], which allows a client Alice to delegate her computation task to a quantum server Bob without revealing any information about the inputs, outputs and computation to Bob. In this protocol, we assume that Alice is able to prepare single qubits randomly chosen from a finite set and send them to the server. But, other than this she does not need to have any other quantum computational power or quantum memory. Bob has the balance of the required quantum computational resources but may not be trusted by Alice. After Bob received the qubits prepared by Alice, they proceed with a two-way classical communication that allows Alice to drive the computation by giving single-qubit measurement instructions (which depend on previous measurement outcomes) to Bob. At the end, Alice can deduce the desired computational results but

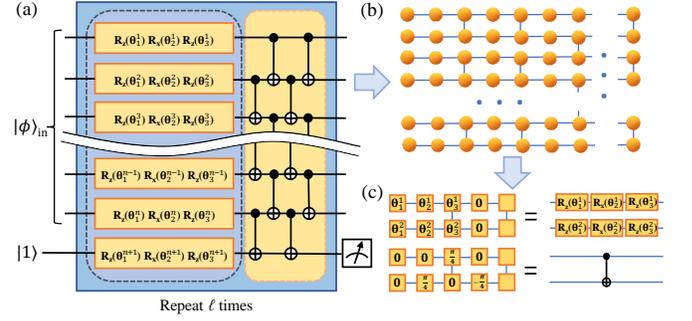


FIG. 2. (a) The structure of the variational quantum classifiers that conveniently fits the protocol of universal blind quantum computation (UBQC) in [23]. This basic structure contains a rotation layer with z - and x -rotations and an entangling layer with controlled-NOT gates, and it will repeat ℓ times to form the desired classifier. (b) the brickwork state used in the UBQC protocol. (c) The implementations of the rotations and controlled-NOT gates by measuring the brickwork state along specific directions. See [43] for details.

Bob learns nothing about Alice's inputs, outputs, and the computation. This protocol is unconditional secure, in the sense that it does not rely on any computational assumptions and holds regardless of what actions a cheating Bob undertakes.

For our purpose, we take advantage of the security and integrity provided by the UBQC protocol and consider delegating quantum machine learning tasks to an untrusted quantum server. For simplicity, we focus on the classification tasks and consider a scenario where N clients want to train collaboratively a shared quantum classifier with their private data unrevealed. To this end, each client can delegate the computation to the server by using the UBQC protocol [23]. For the training process, we define the loss function as the cross entropy

$$L(h(|\psi\rangle_{in}; \Theta), \mathbf{a}) = - \sum_k a_k \log g_k, \quad (1)$$

where $\mathbf{a} \equiv (a_1, \dots, a_m)$ is the label corresponding to the input state $|\phi\rangle_{in}$ in the form of one-hot encoding, h denotes the hypothesis function determined by the variational quantum classifier (with parameters collectively denoted by Θ), and $\mathbf{g} \equiv (g_1, \dots, g_m) = \text{diag}(\rho_{out})$ gives all the probabilities of the corresponding classification categories with ρ_{out} denoting the output state of the classifier [44]. To minimize the loss function, the "parameter shift rule" [45–47] can be used as a subroutine to compute the gradients (see the Supplementary Material [43] for details). When $N = 1$, we only have a single client and we can assume that the quantum classifier is kept in confidential. In this case, the client Alice can use the UBQC protocol directly and her privacy is preserved perfectly. Indeed, the server even does not know that Alice is dealing with a machine learning task. Whereas, when N is larger than one multiple clients share the same quantum classifier. In this case, it is more reasonable to assume that the quantum classifier and its update information are publicly announced. As a result, an adversary may exploit these information to retrieve the clients' private data, through for instance

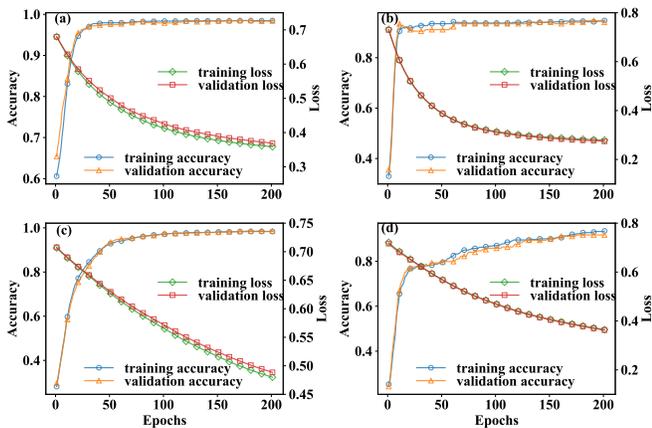


FIG. 3. Numerical results on the demonstration of private quantum distributed learning based on the universal blind quantum computing protocol introduced in [23]. (a) and (b) show the results for the single-client case with the MNIST dataset and the Wisconsin Diagnostic Breast Cancer dataset, respectively. Similarly, (c) and (d) show the results for the multi-client case for the same two datasets, with differential privacy adapted. See [43] for details.

the gradient attack method [48]. To better protect the privacy for each client, we should incorporate the differential privacy method to our protocol, as discussed below.

Single-party delegated learning.—We first consider the $N = 1$ case. Suppose Alice has a variational quantum classifier with structures shown in Fig. 2 (a). This classifier can be decomposed into rotation and entangling layers with rotation and the controlled-NOT gates that are conveniently implemented by measuring each site of the brickwork state along different angles, as illustrated in Fig. 2 (b,c). Alice wants to finish a classification task with this quantum classifier, but she does not have a quantum computer and thus needs to delegate the task to Bob who maintains a quantum server. As mentioned in the above discussion, she can accomplish this through applying the UBQC protocol directly. For concreteness, here we consider two classification tasks, one about identifying handwritten digit images "1 or 9" in the MNIST dataset [49] and the other about classifying data samples in the Wisconsin Diagnostic Breast Cancer (WDBC) dataset [50]. We numerically simulate the training and validation process [43] and our results are plotted in Fig. 3 (a, b). From these figures, we find that both the training and validation accuracies increase rapidly at the beginning of the training process and then saturate roughly at 98% and 94% for the MNIST and WDBC datasets, respectively. This shows that the delegated learning indeed can achieve a decent performance on these classification tasks. In addition, considering the unavoidable imperfections of quantum devices in reality, the preparation of the single qubits and the transfer of these qubits on the Alice's side and the measurement on the Bob's side will inevitably bring some noises. To benchmark the robustness of this protocol, we also carry out numerical simulations with various noises and different noise strengths. Our detailed results are

Algorithm 1 Quantum private distributed learning

Input The untrained model h with parameters Θ , the loss function L , the number of clients N , the number of samples belongs to one client n , the batch size n_b , the number of iterations T , the learning rate ϵ , the Adam optimizer f_{Adam} , the gradient norm bound R , the Laplace noise \mathcal{N} , the noise strength μ

Output The trained model

Initialization: randomly generate a length- T string S with each element corresponding to an index of a client

for $i \in [T]$ **do**

Choose the client with index S_i and randomly choose n_b samples among the n samples of the client

Calculate the gradient under the UBQC protocol $\vec{g}_i \leftarrow \frac{1}{n_b} \sum_{k=1}^{n_b} \nabla L(h(|\psi\rangle_{i,k}; \Theta), \mathbf{a}_{i,k})$

Clip the gradient: $\vec{g}_i \leftarrow \vec{g}_i / \max(1, \frac{\|\vec{g}_i\|}{R})$

Add noise: $\tilde{g}_i \leftarrow \vec{g}_i + \frac{2R}{\mu} \mathcal{N}(0, 1)$

Updates: $\Theta \leftarrow \Theta - \epsilon \cdot f_{\text{Adam}}(\tilde{g}_i)$

end for

Output the trained model

shown in the Supplementary Materials [43].

We remark that we deliberately choose the variational quantum classifier above because it is "native" to measurement-based quantum computing and can be conveniently implemented when applying the UBQC protocol. One may choose other variational quantum classifiers, such as the hierarchical quantum classifiers [51] or quantum convolutional neural networks [52]. Yet, for these classifiers some additional modifications might be needed so as to match the particular brickwork state used in the UBQC protocol. In addition, we stress that in this single-client scenario Alice can keep the structure of her quantum classifier in confidential, and Bob learns nothing about Alice's inputs, outputs, or desired computation. Alice's privacy can be perfectly preserved.

Multi-client private distributed learning.—We now turn to the case of $N > 1$, where several clients collaboratively train a shared model with their private data unrevealed. One may regard this as the quantum counterpart of classical federated learning [34, 35]. Assume that each client has n training samples and the shared quantum classifier has the same structure as that in Fig. 2 (a). In every turn of the collaborative training, one client will use the UBQC protocol to delegate the computation of the gradient to the server with the private data the client holds. Then the gradients will be uploaded to update the parameters of the shared quantum classifier with the following rule:

$$\Theta_{t+1} = \Theta_t - \epsilon \cdot f(\vec{g}), \quad (2)$$

where Θ_t denotes collectively the parameters at the t -th step, ϵ is the learning rate, \vec{g} is the vectorized gradient, and $f(\vec{g})$ is a modification function of the gradient which we take the Adam optimizer [53] in this paper. However, if there are eavesdroppers with access to the model, uploading the gradients directly can be attacked by reverse engineering to recover the input

Algorithm 2 Gradient attack algorithm

Input The target gradient \vec{g}_t , the model h with parameters Θ , the loss function L for training the model, the string S that contains n different labels, the number of iterations T , the learning rate ϵ , the Adam optimizer f_{Adam}

Output The recovered input data and label

Prepare the initial state as a uniform quantum state $|\phi\rangle \leftarrow |\phi_0\rangle$

for $\mathbf{k} \in S$ **do**

Define the loss function for the gradient attack as $L_{\text{GA}}(|\phi\rangle) \leftarrow \|\nabla L(h(|\phi\rangle; \Theta), \mathbf{k}) - \vec{g}_t\|^2$

for $i \in [T]$ **do**

Calculate the gradient $|\vec{g}_i\rangle \leftarrow \nabla_{\text{in}} L_{\text{GA}}(|\phi\rangle)$

Updates: $|\phi\rangle \leftarrow |\phi\rangle - \epsilon \cdot f_{\text{Adam}}(|\vec{g}_i\rangle)$

Normalization: $|\phi\rangle \leftarrow \frac{1}{\| |\phi\rangle \|} |\phi\rangle$

end for

If the final loss function with this label \mathbf{k} is the minimum, the final state $|\phi\rangle$ and label \mathbf{k} will be set as the recovered input data and label

end for

Output the recovered input data and label

corresponding to the gradients [48]. To handle this problem, we adapt the idea of differential privacy [36–38] to our protocol and add appropriate noise like Laplace noise or Gaussian noise to the parameters before uploading them. The feasibility can be understood as follows. Before adding noise, there might be an input whose gradient exactly corresponds to the uploaded gradient, thus the real input can be recovered by some optimization process. With adaption of differential privacy, the attacker only sees a combination of the real gradient and random noise. The result of the attack becomes random and it is hard for the attacker to differentiate the gradient corresponding to the recovered input and the real gradient obtained from the original data. Yet, it is also worthwhile to point out that in this case the clients are training a shared model which is public. So if the updated parameters from a client can make the model behave better, it is inevitable that some information about this client’s data will be leaked. This is in sharp contrast to the single-client case, where the privacy can be perfect as discussed above. In real experiments, the intrinsic quantum noises may offer the randomness the protocol needs and be beneficial for privacy protection in these tasks [54]. With this strategy, the clients can iteratively train the shared model to find the optimal parameters. The pseudocode for our protocol is given in Algorithm 1. To benchmark how our protocol works, we also carry out numerical simulations on the MNIST and WDBC datasets. Our results are shown in Fig. 3(c) and Fig. 3(d), from which it is clear that the increase of the accuracies are slower than that for the single-client case. This is due to the adaption of differential privacy, which will downgrade the accuracy in general. To account for experimental imperfections, we also carry out numerical simulations with different noise strengths and our results are shown in the Supplementary Materials [43].

Gradient attack.—So far we have presented a concrete pro-

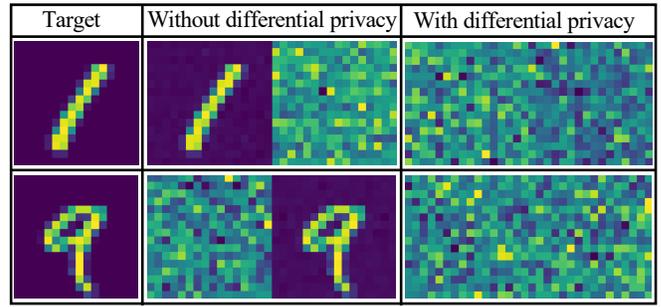


FIG. 4. The performance of the gradient attack when trying to recover the input data based on the updates the client uploaded. Here, we simulate the attacks numerically with two handwritten digit images of "1" and "9". For each image, the attacker obtains two recovered images, rather than one image since he does not know the true label and needs to try both labels, through this gradient attack method. Without differential privacy, one of the recovered images has very high fidelity with the original image, thus the private data information is leaked. Whereas with differential privacy, both the recovered images seem random and the private data of the client is indeed preserved.

ocol for private distributed learning with the method of differential privacy to prevent malicious attacks. To better illustrate why differential privacy is needed, we now apply the gradient attack [48] to test the performance before and after adding noise to the gradients. Generally, when training a variational quantum circuit to minimize a certain loss function, we will calculate the derivative of this loss function with respect to the model parameters. However, if an eavesdropper wants to attack a client’s data, what he has is the updated parameters. So inversely, the task becomes recovering the input of the circuit with the updates. We first demonstrate how the gradient attack works in the situation where we upload the gradients directly without adding noises. We take the two handwritten digits with labels "1" and "9" respectively as a client’s input to be attacked and suppose the attack happens at step t with model parameters Θ . When the client uses the real input $|\phi_r\rangle$ to do the training, the gradient descent method will give the target gradient $\vec{g}_t = \nabla L(h(|\phi_r\rangle; \Theta), \mathbf{a})$. On the other hand, since the eavesdropper does not know any information about the input data as promised by the UBQC protocol, he may start with a uniform state $|\phi\rangle = |\phi_0\rangle$. Now the loss function for gradient attack is defined as

$$L_{\text{GA}}(|\phi\rangle) = \|\nabla L(h(|\phi\rangle; \Theta), \mathbf{a}) - \nabla L(h(|\phi_r\rangle; \Theta), \mathbf{a})\|^2,$$

which measures the distance between the current gradient to the target. We mention that the eavesdropper has no information about the label \mathbf{a} which is contained in the loss function, so the attack needs to be carried out with all the possible labels. To minimize this loss function, we need to compute the derivatives of it with respect to the input samples rather than the circuit parameters, and the concrete algorithm is given in Algorithm 2. The numerical demonstration for this algorithm to attack the handwritten digits without differential privacy is shown in the second column of Fig. 4, where the input digits

are recovered with very high fidelity. However, if we clip the gradients with a bound R and add appropriate Laplace noise to the uploaded gradients which is taken by Algorithm 1 as a subroutine and will slightly reduce the training accuracy, the loss function becomes

$$L_{\text{GA}}(|\phi\rangle) = \|\nabla L(h(|\phi\rangle; \Theta), \mathbf{a}) - \mathcal{N}(\vec{g}_t)\|^2, \quad (3)$$

where \mathcal{N} contains the operations of adding noise and clipping the gradients. It turns out that the recovered digits from the noisy gradients have a clear difference from the real ones as shown in the third column of Fig. 4. It should be noted that even though the noise can help with the privacy in the figures, the attacker may also get some information about the labels of the input according to the curve of the training process, i.e., the loss function with the right label may achieve the smallest value. One way to strengthen the privacy is to compute the gradients with several samples with different labels at a step, e.g. stochastic gradient descent with the size of each batch larger than one, and take their average value as the updates.

Discussion and conclusion.—Although we focus on classification problems in this paper, our protocol carries over straightforwardly to other quantum learning problems as well. We stress that, compared with its classical counterparts, our quantum private distributed learning protocol may achieve exponential speedup for certain learning tasks. For instance, in Ref. [18] a quantum machine learning algorithm based on generative models has been introduced and rigorously proved to have exponential speedup over any classical algorithms for some instances if a quantum computer cannot be efficiently simulated classically. By using our protocol, a client may also delegate the training of this generative model to a quantum server while maintaining perfect privacy. For these instances with no efficient classical algorithm, our protocol may achieve both exponential speedup and privacy protection. In addition, our protocol requires that the clients have the capability of preparing single qubits, which may not be satisfied under certain circumstances. To release this requirement, one could exploit other blind quantum computing protocols concerning purely classical clients [23, 55, 56]. Given the fact that privacy protection becomes more and more important in the data-intensive and sensitive society, an experimental demonstration of quantum private distributed learning would be a crucial step toward practical applications of quantum technologies for tackling security related problems in artificial intelligence in the future.

In summary, we have introduced a protocol for quantum private distributed learning based on blind quantum computing, where several clients with inadequate data can collaboratively train a quantum learning model while keeping their sensitive data unrevealed. To benchmark the effectiveness of our protocol, we carried out extensive numerical simulations with different real-life datasets and encoding strategies, taking into consideration of experimental imperfections. In addition, we also demonstrated that our protocol survives the well-known gradient attack after the adaption of differential privacy. Our results show the intriguing potential of achieving

large-scale private distributed learning with both near-term and future quantum technologies, which provide a valuable guide for exploring quantum advantages in real-life machine learning applications from the security perspective.

We thank Haoyu Guo, Wenjie Jiang, Zhide Lu, Peixin Shen for helpful discussions. This work is supported by the start-up fund from Tsinghua University (Grant. No. 53330300320), the National Natural Science Foundation of China (Grant. No. 12075128), and the Shanghai Qi Zhi Institute.

* These authors contributed equally to this work.

† dldeng@tsinghua.edu.cn

- [1] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, Quantum machine learning, *Nature* **549**, 195 (2017).
- [2] V. Dunjko and H. J. Briegel, Machine learning & artificial intelligence in the quantum domain: a review of recent progress, *Rep. Prog. Phys.* **81**, 074001 (2018).
- [3] S. D. Sarma, D.-L. Deng, and L.-M. Duan, Machine learning meets quantum physics, *Phys. Today* **72**, 48 (2019).
- [4] S. J. Russell and P. Norvig, *Artificial intelligence: a modern approach* (2010).
- [5] Y. LeCun, Y. Bengio, and G. Hinton, Deep learning, *Nature* **521**, 436 (2015).
- [6] M. I. Jordan and T. M. Mitchell, Machine learning: Trends, perspectives, and prospects, *Science* **349**, 255 (2015).
- [7] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. Van Den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, *et al.*, Mastering the game of go with deep neural networks and tree search, *Nature* **529**, 484 (2016).
- [8] D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton, *et al.*, Mastering the game of go without human knowledge, *Nature* **550**, 354 (2017).
- [9] A. W. Senior, R. Evans, J. Jumper, J. Kirkpatrick, L. Sifre, T. Green, C. Qin, A. Židek, A. W. Nelson, A. Bridgland, *et al.*, Improved protein structure prediction using potentials from deep learning, *Nature* **577**, 706 (2020).
- [10] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, and L. Zdeborová, Machine learning and the physical sciences, *Rev. Mod. Phys.* **91**, 045002 (2019).
- [11] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. Brandao, D. A. Buell, *et al.*, Quantum supremacy using a programmable superconducting processor, *Nature* **574**, 505 (2019).
- [12] H.-S. Zhong, H. Wang, Y.-H. Deng, M.-C. Chen, L.-C. Peng, Y.-H. Luo, J. Qin, D. Wu, X. Ding, Y. Hu, *et al.*, Quantum computational advantage using photons, *Science* **370**, 1460 (2020).
- [13] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2010).
- [14] A. W. Harrow, A. Hassidim, and S. Lloyd, Quantum algorithm for linear systems of equations, *Phys. Rev. Lett.* **103**, 150502 (2009).
- [15] S. Lloyd, M. Mohseni, and P. Rebentrost, Quantum principal component analysis, *Nat. Phys.* **10**, 631 (2014).
- [16] V. Havlíček, A. D. Córcoles, K. Temme, A. W. Harrow, A. Kandala, J. M. Chow, and J. M. Gambetta, Supervised learning with

- quantum-enhanced feature spaces, *Nature* **567**, 209 (2019).
- [17] M. Schuld and N. Killoran, Quantum machine learning in feature hilbert spaces, *Phys. Rev. Lett.* **122**, 040504 (2019).
- [18] X. Gao, Z.-Y. Zhang, and L.-M. Duan, A quantum machine learning algorithm based on generative models, *Sci. Adv.* **4**, eaat9004 (2018).
- [19] S. Lloyd and C. Weedbrook, Quantum generative adversarial learning, *Phys. Rev. Lett.* **121**, 040502 (2018).
- [20] L. Hu, S.-H. Wu, W. Cai, Y. Ma, X. Mu, Y. Xu, H. Wang, Y. Song, D.-L. Deng, C.-L. Zou, *et al.*, Quantum generative adversarial learning in a superconducting quantum circuit, *Sci. Adv.* **5**, eaav2761 (2019).
- [21] P. Rebentrost, M. Mohseni, and S. Lloyd, Quantum support vector machine for big data classification, *Phys. Rev. Lett.* **113**, 130503 (2014).
- [22] S. Barz, E. Kashefi, A. Broadbent, J. F. Fitzsimons, A. Zeilinger, and P. Walther, Demonstration of blind quantum computing, *Science* **335**, 303 (2012).
- [23] A. Broadbent, J. Fitzsimons, and E. Kashefi, Universal blind quantum computation, in *2009 50th Annual IEEE Symposium on Foundations of Computer Science* (IEEE, 2009) pp. 517–526.
- [24] A. M. Childs, Secure assisted quantum computation, [arXiv: quant-ph/0111046](https://arxiv.org/abs/quant-ph/0111046) (2001).
- [25] J. F. Fitzsimons and E. Kashefi, Unconditionally verifiable blind quantum computation, *Phys. Rev. A* **96**, 012303 (2017).
- [26] J. F. Fitzsimons, Private quantum computation: an introduction to blind quantum computing and related protocols, *npj Quantum Inf.* **3**, 1 (2017).
- [27] V. Giovannetti, L. Maccone, T. Morimae, and T. G. Rudolph, Efficient universal blind quantum computation, *Phys. Rev. Lett.* **111**, 230501 (2013).
- [28] T. Morimae and K. Fujii, Blind quantum computation protocol in which alice only makes measurements, *Phys. Rev. A* **87**, 050301 (2013).
- [29] T. Morimae and K. Fujii, Blind topological measurement-based quantum computation, *Nat. Commun.* **3**, 1 (2012).
- [30] T. Sueki, T. Koshihara, and T. Morimae, Ancilla-driven universal blind quantum computation, *Phys. Rev. A* **87**, 060301 (2013).
- [31] F. Zerka, S. Barakat, S. Walsh, M. Bogowicz, R. T. Leijenaar, A. Jochems, B. Miraglio, D. Townsend, and P. Lambin, Systematic review of privacy-preserving distributed machine learning from federated databases in health care, *JCO clinical cancer informatics* **4**, 184 (2020).
- [32] A. Hosny, C. Parmar, J. Quackenbush, L. H. Schwartz, and H. J. Aerts, Artificial intelligence in radiology, *Nature Reviews Cancer* **18**, 500 (2018).
- [33] M. X. Chen, B. N. Lee, G. Bansal, Y. Cao, S. Zhang, J. Lu, J. Tsay, Y. Wang, A. M. Dai, Z. Chen, *et al.*, Gmail smart compose: Real-time assisted writing, in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (2019) pp. 2287–2295.
- [34] J. Konečný, H. B. McMahan, F. X. Yu, P. Richtárik, A. T. Suresh, and D. Bacon, Federated learning: Strategies for improving communication efficiency, [arXiv:1610.05492](https://arxiv.org/abs/1610.05492) (2016).
- [35] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, Communication-efficient learning of deep networks from decentralized data, in *Artificial Intelligence and Statistics* (PMLR, 2017) pp. 1273–1282.
- [36] C. Dwork, Differential privacy: A survey of results, in *International conference on theory and applications of models of computation* (Springer, 2008) pp. 1–19.
- [37] C. Dwork, A. Roth, *et al.*, The algorithmic foundations of differential privacy., *Foundations and Trends in Theoretical Computer Science* **9**, 211 (2014).
- [38] M. Abadi, A. Chu, I. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar, and L. Zhang, Deep learning with differential privacy, in *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security* (2016) pp. 308–318.
- [39] Y. Huang, Z. Song, K. Li, and S. Arora, Instahide: Instance-hiding schemes for private distributed learning, in *International Conference on Machine Learning* (PMLR, 2020) pp. 4507–4518.
- [40] R. Bassily, A. Smith, and A. Thakurta, Private empirical risk minimization: Efficient algorithms and tight error bounds, in *2014 IEEE 55th Annual Symposium on Foundations of Computer Science* (IEEE, 2014) pp. 464–473.
- [41] A. Beimel, H. Brenner, S. P. Kasiviswanathan, and K. Nissim, Bounds on the sample complexity for private learning and private data release, *Machine learning* **94**, 401 (2014).
- [42] C. Gentry, Fully homomorphic encryption using ideal lattices, in *Proceedings of the forty-first annual ACM symposium on Theory of computing* (2009) pp. 169–178.
- [43] See Supplemental Material at [URL will be inserted by publisher] for the sketch of the universal blind quantum computation, the details of the numerical simulations, the robustness test, differential privacy with different noise strengths, and different encoding strategies.
- [44] S. Lu, L.-M. Duan, and D.-L. Deng, Quantum adversarial machine learning, *Phys. Rev. Research* **2**, 033212 (2020).
- [45] K. Mitarai, M. Negoro, M. Kitagawa, and K. Fujii, Quantum circuit learning, *Phys. Rev. A* **98**, 032309 (2018).
- [46] J. Li, X. Yang, X. Peng, and C.-P. Sun, Hybrid quantum-classical approach to quantum optimal control, *Phys. Rev. Lett.* **118**, 150503 (2017).
- [47] M. Schuld, V. Bergholm, C. Gogolin, J. Izaac, and N. Killoran, Evaluating analytic gradients on quantum hardware, *Phys. Rev. A* **99**, 032331 (2019).
- [48] L. Zhu, Z. Liu, and S. Han, Deep leakage from gradients, in *Advances in Neural Information Processing Systems* (2019) pp. 14774–14784.
- [49] Y. LeCun, C. Cortes, and C. Burges, *Mnist handwritten digit database* (1998).
- [50] W. H. Wolberg, N. Street, and O. L. Mangasarian, *UCI Machine Learning Repository: Breast Cancer Wisconsin (Diagnostic) Data Set* (1992).
- [51] E. Grant, M. Benedetti, S. Cao, A. Hallam, J. Lockhart, V. Stojevic, A. G. Green, and S. Severini, Hierarchical quantum classifiers, *npj Quantum Inf.* **4**, 1 (2018).
- [52] I. Cong, S. Choi, and M. D. Lukin, Quantum convolutional neural networks, *Nat. Phys.* **15**, 1273 (2019).
- [53] D. P. Kingma and J. Ba, Adam: A method for stochastic optimization, [arXiv:1412.6980](https://arxiv.org/abs/1412.6980) (2014).
- [54] Y. Du, M.-H. Hsieh, T. Liu, D. Tao, and N. Liu, Quantum noise protects quantum classifiers against adversaries, [arXiv:2003.09416](https://arxiv.org/abs/2003.09416) (2020).
- [55] H.-L. Huang, Q. Zhao, X. Ma, C. Liu, Z.-E. Su, X.-L. Wang, L. Li, N.-L. Liu, B. C. Sanders, C.-Y. Lu, *et al.*, Experimental blind quantum computing for a classical client, *Phys Rev Lett* **119**, 050503 (2017).
- [56] A. Mantri, T. F. Demarie, N. C. Menicucci, and J. F. Fitzsimons, Flow ambiguity: A path towards classically driven blind quantum computation, *Phys. Rev. X* **7**, 031004 (2017).
- [57] V. Bergholm, J. J. Vartiainen, M. Möttönen, and M. M. Salomaa, Quantum circuits with uniformly controlled one-qubit gates, *Phys. Rev. A* **71**, 052330 (2005).
- [58] M. Plesch and Č. Brukner, Quantum-state preparation with universal gate decompositions, *Phys. Rev. A* **83**, 032302 (2011).

Supplementary Material: Private Distributed Learning Through Blind Quantum Computing

In this Supplementary Material, we firstly give a review of the UBQC from [23] and illustrate how to adapt our learning tasks to the UBQC protocol. Then we will give the detailed settings of our numerical experiments in the following. As mentioned in the main text, we will firstly test the robustness of the quantum classifiers with different noise strengths. The numerical simulations of differential privacy with different noise strengths will be presented as well. In the last part, we will test a different data encoding strategy that encodes the data in the rotation angles through the quantum circuit and give some discussions about it.

I. REVIEW AND USAGE OF THE UNIVERSAL BLIND QUANTUM COMPUTATION

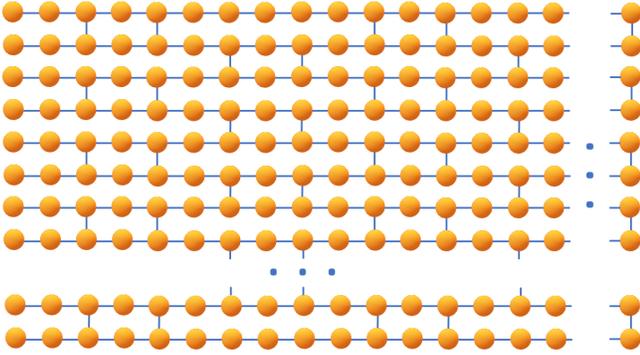


FIG. S1. The structure for a universal blind quantum computation scheme on the server's side adapted from [23]. Every orange ball denotes a qubit received from Alice, and the connected lines denote controlled-Z gates.

The goal of blind quantum computation is to delegate a computational task to a quantum server while keeping the inputs, outputs, and computation details private. Suppose there is a client Alice and a quantum server Bob. The universal blind quantum computation from [23] can be achieved where Alice only needs the quantum power to prepare single qubits and sends them to Bob. To start with, the structure on the Bob's side which is called the brickwork state is shown in Fig. S1. For each qubit in this structure, Alice prepares a state in $\frac{1}{\sqrt{2}}(|0\rangle + e^{i\theta}|1\rangle)$ where θ is a uniformly random number from a finite or infinite set and sends it to the corresponding spot. After all the qubit preparations, Bob will apply controlled-Z gates to all the pairs connected by lines to create an entangled state. Then the final state is suitable for measurement based quantum computations and Alice can delegate the computation by two-way classical communications.

Suppose there are n columns and m rows in the brickwork structure. Let $\phi_{x,y}$ be the measurement angle on column x , row y to achieve the computation given the condition that the

initial sent qubits are all in state $\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$. Alice will have these angles in mind and give the measurement angle $\delta_{x,y} = \phi_{x,y} + \theta_{x,y} + \pi r_{x,y}$ to Bob where $\theta_{x,y}$ is the random angle for point (x, y) in the single qubit preparations and $r_{x,y}$ is a uniformly random binary number. It can be proved that in this way, correctness and blindness can both be achieved. For the correctness, the measurement on the state $\frac{1}{\sqrt{2}}(|0\rangle + e^{i(\phi+\theta)}|1\rangle)$ in the basis $\frac{1}{\sqrt{2}}(|0\rangle \pm e^{i(\psi+\theta)}|1\rangle)$ is the same as the measurement on the state $\frac{1}{\sqrt{2}}(|0\rangle + e^{i\phi}|1\rangle)$ in the basis $\frac{1}{\sqrt{2}}(|0\rangle \pm e^{i\psi}|1\rangle)$, thus leaving no influence on the original computation. As for the random binary number $r_{x,y}$, if it is zero, the measurement angle is not affected. If it is one, Alice can simply flip the measurement result to get the correct answer. For the blindness, the information that Bob has can be divided into three parts: (1) the single qubits sent from Alice; (2) the classical information from Alice to apply the measurement based computation; (3) the binary information from the measurement outcomes. Firstly, the uniformly random number $\theta_{x,y}$ can make the information encoded in the qubit and the classical information sent by Alice uniformly random on Bob's side. Secondly, the binary measurement outcomes will appear uniformly random on Bob's side as promised by $r_{x,y}$. Furthermore, an authentication protocol to detect a cheating server has also been proposed in [23] in details. Thus the feasibility of this protocol has a theoretical guarantee.

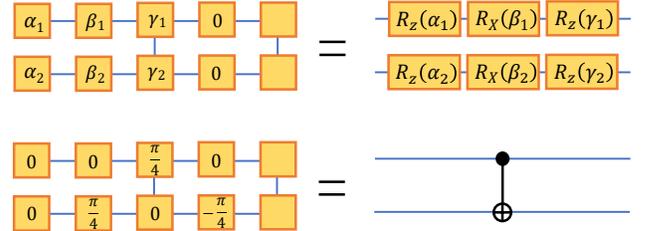


FIG. S2. The construction of the variational quantum classifier in our work including the construction of single-qubit gates (up) and controlled-NOT gates (down) adapted from [23], where the corrections that consider the measurement outcomes are not explicitly shown.

Since this protocol is universal for quantum computation, we can construct a variety of circuits for different uses. In our work, we take the advantage that it is efficient to construct single-qubit gates and controlled-NOT gates in this framework as illustrated in Fig. S2 to build a variational quantum classifier as already shown in Fig. 2. In Fig. S2, the angles on the left side denotes the measurement angles to do the measurement based computation, where the corrections that consider the measurement outcomes are not explicitly shown. Then the achieved single-qubit gates and controlled-NOT gates are presented on the right side. In this setting, the parameters encoded in the single-qubit gates are continuous and the controlled-NOT gates are fixed. So in order to encrypt the continuous information, the $\theta_{x,y}$ should be sam-

pled from a continuous set such as $R[0, 2\pi]$ we adopt.

II. DETAILS OF OUR NUMERICAL EXPERIMENTS

Data encoding.—In the main text, the data of handwritten digits and breast cancer are encoded as the input of the circuit with the method of amplitude encoding, which enables us to represent $N = 2^n$ dimensional classical data with n qubits. During the private delegated computing process, Alice can only prepare single qubit states and send them to Bob. So it points out the need to construct highly-entangled target states on Bob's side. According to the work from [57, 58], we can decompose any quantum state into a series of single-qubit rotations and two-qubit gates like controlled-NOT gates. The single-qubit rotations can be directly created in the structure of UBQC. As for the controlled-NOT gates, if the two qubits are adjacent, the controlled-NOT gates can be directly constructed as shown before. Otherwise if the two qubits are not adjacent, we can use SWAP gates and controlled-NOT gates to accomplish the target controlled-NOT gate where the SWAP gates can also be constructed using three controlled-NOT gates. In this way, we can use the protocol above as a subroutine to achieve the amplitude encoding strategy.

Gradient descent.—In the optimization process, we set the loss function L as cross entropy which takes the expectation value of the ancillary qubits in the standard basis as a component. The output state of the ancillary qubits can be written as ρ_{out} . Let $\mathbf{g} \equiv (g_1, \dots, g_{2^m}) = \text{diag}(\rho_{out})$ and \mathbf{a} be the label corresponding to the input state $|\phi\rangle_{in}$ in the form of one-hot encoding. Then we can formalize the loss function as

$$L(h(|\psi\rangle_{in}; \Theta), \mathbf{a}) = - \sum_k a_k \log g_k.$$

It can be seen that reducing the loss function is consistent with reducing the distance between \mathbf{a} and \mathbf{g} . To reduce the loss function, computing the derivatives of L with respect to the circuit parameters can be transformed into computing the derivatives of some expectation values with respect to these circuit parameters according to the chain rule. That is,

$$\frac{\partial L(h(|\psi\rangle_{in}; \Theta), \mathbf{a})}{\partial \theta} = - \sum_k \frac{a_k}{g_k} \frac{\partial g_k}{\partial \theta}.$$

The next step that computes the derivatives of g_k to the circuit parameters can be done with the "parameter shift rule" since g_k can be seen as an expectation value of an observable which we denote as B_k here [45–47]:

$$\frac{\partial g_k}{\partial \theta} = \frac{\partial \langle B_k \rangle}{\partial \theta} = \frac{\langle B_k \rangle^+ - \langle B_k \rangle^-}{2},$$

where $\langle B_k \rangle^\pm$ denotes the expectation value of B_k with the parameter θ being $\theta \pm \frac{\pi}{2}$. With these preparations, we can optimize the variational quantum circuit with gradient based methods.

Numerical settings.—In the main text, we present the numerical simulations for single-party delegated learning and multiparty distributed learning with two datasets. Here in Table S1 and Table S2 we list some numerical settings for these experiments.

	mnist classification	tumour diagnosis
classes	2("1" or "9")	2
circuit depth	30	30
num of qubits	8	6
learning rate	0.001	0.001
num of iterations	300	300
batch size	50	100
optimizer	Adam	Adam

TABLE S1. Parameter settings for single-party delegated learning.

	mnist classification	tumour diagnosis
classes	2("1" or "9")	2
gradient bound	0.01	0.01
circuit depth	30	30
num of qubits	8	6
learning rate	0.001	0.001
num of iterations	1500	1000
num of clients	10	4
training set(per client)	200	100
batch size(per client)	50	50
optimizer	Adam	Adam

TABLE S2. Parameter settings for multiparty distributed learning.

III. ROBUSTNESS TEST FOR THE VARIATIONAL QUANTUM CLASSIFIERS

Considering the inevitable noise in the real quantum devices, both the qubits sent by Alice and the computation executed by Bob may deviate from the noiseless situation. To explore how well the classifier we have proposed performs against the random noise, we carry out several numerical simulations with noise included.

We divide the noise into two parts: (1) the noise in the input training data, which may result from the qubits sent by Alice and the measurements to generate the entangled state by Bob; (2) the noise in the parameterized circuit, which may result from the qubits sent by Alice and the measurements to do the computation by Bob. For the first part, we add noise, which is a number normally distributed with mean value zero and standard bias one times the input noise strength that we can set with different values, to every element of the input vector. For the second part, we add noise, which is 0.001 times a number normally distributed with mean value zero and standard bias one, to every element of the gradient independently during the training process. The training process for two datasets are shown in Fig. S3 and Fig. S4, where the model can still achieve a decent performance when the noise strength is not

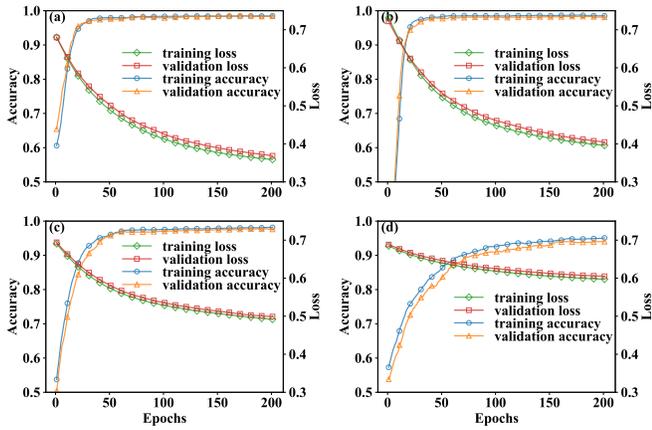


FIG. S3. The robustness test with the handwritten digits from the MNIST dataset: (a) The zero noise situation corresponding to Fig. 3(a). (b) The noise strength of the input state is set as 0.02. (c) The noise strength of the input state is set as 0.05. (d) The noise strength of the input state is set as 0.1.

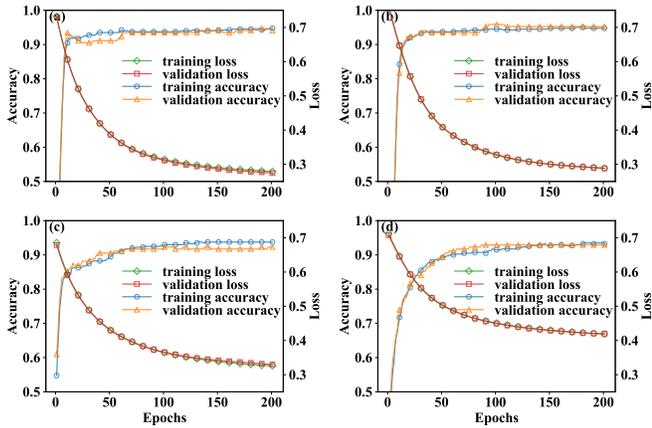


FIG. S4. The robustness test with the data from the Wisconsin Diagnostic Breast Cancer dataset: (a) The zero noise situation corresponding to Fig. 3(b). (b) The noise strength of the input state is set as 0.02. (c) The noise strength of the input state is set as 0.05. (d) The noise strength of the input state is set as 0.1.

too large. For concreteness, in both the MNIST dataset and the Wisconsin Diagnostic Breast Cancer dataset, the variational quantum learning model shows good resilience to noise when the noise strength is below 0.02. With the increase of the noise, the accuracy will drop and it will take more time to converge.

IV. DIFFERENTIAL PRIVACY WITH DIFFERENT NOISE STRENGTHS

The noise in the differential privacy plays a beneficial role to protect the privacy, which is different from the noise in the robustness test section which may harm the training. In this section, we give out the numerical simulations for the differential privacy scheme with different noise strengths. The ba-

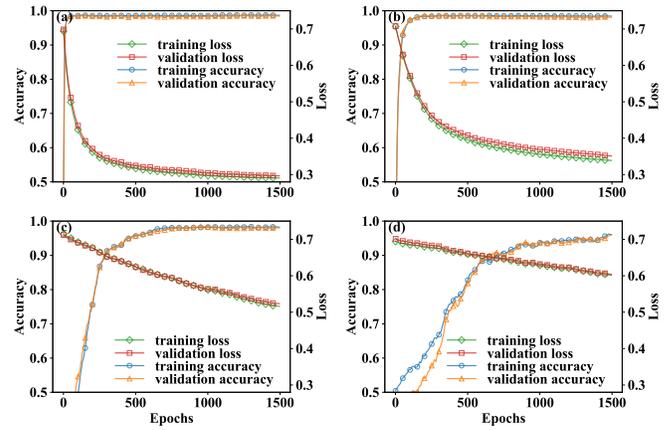


FIG. S5. The differential privacy with the handwritten digits from the MNIST dataset: (a) The zero noise situation $\mu = +\infty$. (b) The noise strength $\mu = 1$. (c) The noise strength $\mu = 0.1$. (d) The noise strength $\mu = 0.05$. There is more noise when μ is smaller as defined in Algorithm 1.

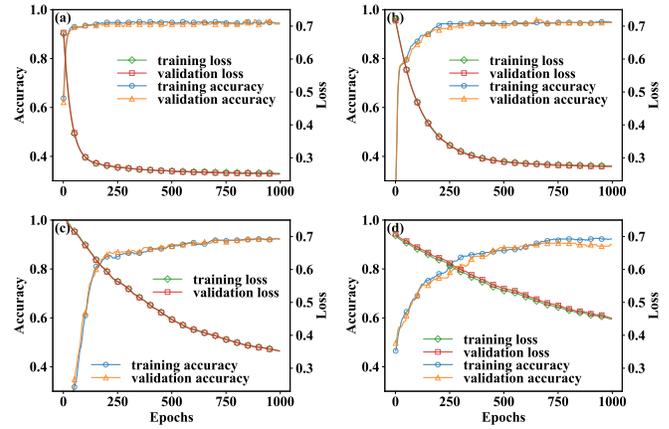


FIG. S6. The differential privacy with the data from the Wisconsin Diagnostic Breast Cancer dataset: (a) The zero noise situation $\mu = +\infty$. (b) The noise strength $\mu = 1$. (c) The noise strength $\mu = 0.2$. (d) The noise strength $\mu = 0.1$. There is more noise when μ is smaller as defined in Algorithm 1.

sic settings have already been given in Table S2. For the noise strength μ defined in Algorithm 1, we also set four different values, where $\mu = +\infty$ means no noise. The training process for two datasets are shown in Fig. S5 and Fig. S6.

It is apparent from the training curves that when we add more noise, the training accuracy curves will decrease and it will take more steps to converge for both the two datasets. When the noise strength μ is 0.05 in the MNIST dataset and 0.1 in the Wisconsin Diagnostic Breast Cancer dataset, the training becomes very slow and has the trend to fail with increasing noise. Thus the results agree with the point that it is a trade-off between accuracy and privacy.

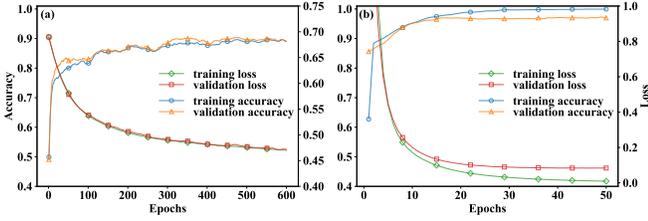


FIG. S7. (a) The training process for the data from MNIST dataset with the encoding strategy that encodes the input features into the circuit parameters. (b) The training process of classical neural networks with two hidden layers to handle the same data as that in (a), which proves that the training data from the circuit encoding process is separable.

V. ENCODING STRATEGY WITH ROTATION ANGLES

In the numerical simulations above, we take the amplitude encoding strategy to encode the input features. Since the client needs to create these entangled states with methods such as the one from [58], it is not efficient compared with the strategy that encodes the features into the circuit parameters. In this section, we will give some brief simulations of the variational quantum classifiers with a different encoding strategy.

Firstly, we still use the parameterized circuit shown in Fig. 2 to encode the input features. The circuit starts with a zero state $|00\dots0\rangle$ as the input. For the encoding process, we use

the framework from Fig. 2 with qubit number 8 and depth 11 which has 272 parameters and initially set these parameters to be 0. Then for the 256-dimensional features from the MNIST dataset, we encode them into the first 256 parameters among the 272 parameters. Then the outputs of the encoding circuit which is a set of 256-dimensional vectors can be used as the inputs for the training circuit and the training can be carried out. To better compare the performance with the amplitude encoding, we take the same settings as the ones from Table S1, and the training curves are given in Fig. S7(a). It can be seen that the trained model can only achieve an accuracy of about 90%, which falls behind the performance with the amplitude encoding strategy. We further demonstrate that, if we extract the outputs of the encoding circuit to feed in a 2-layer classical neural network, the classical model can provide a decent performance as shown in Fig. S7(b).

We use this example to demonstrate that, there are different ways to encode the input features for the training. However, on the one hand, the amplitude encoding can achieve better performance but it requires the inefficient process to prepare the highly entangled states in the delegated tasks. On the other hand, the way we encode the features into the circuit parameters is more efficient but may suffer from the accuracy loss, which calls for better encoding strategies for the future applications. Furthermore, the comparison of the performance of the variational model with the classical model indicates that we need to develop variational quantum learning models with higher expressive power to handle more complex tasks.