Entanglement-Based Machine Learning on a Quantum Computer

X.-D. Cai,[1,2] D. Wu,[1,2] Z.-E. Su,[1,2] M.-C. Chen,[1,2] X.-L. Wang,[1,2] Li Li,[1,2,†]
N.-L. Liu,[1,2,‡] C.-Y. Lu,[1,2,*] and J.-W. Pan[1,2]

[1]*Hefei National Laboratory for Physical Sciences at Microscale and Department of Modern Physics,
University of Science and Technology of China, Hefei, Anhui 230026, China*
[2]*CAS Centre for Excellence and Synergetic Innovation Centre in Quantum Information and Quantum Physics,
University of Science and Technology of China, Hefei, Anhui 230026, China*
(Received 16 November 2014; published 19 March 2015)

Machine learning, a branch of artificial intelligence, learns from previous experience to optimize performance, which is ubiquitous in various fields such as computer sciences, financial analysis, robotics, and bioinformatics. A challenge is that machine learning with the rapidly growing "big data" could become intractable for classical computers. Recently, quantum machine learning algorithms [Lloyd, Mohseni, and Rebentrost, arXiv.1307.0411] were proposed which could offer an exponential speedup over classical algorithms. Here, we report the first experimental entanglement-based classification of two-, four-, and eight-dimensional vectors to different clusters using a small-scale photonic quantum computer, which are then used to implement supervised and unsupervised machine learning. The results demonstrate the working principle of using quantum computers to manipulate and classify high-dimensional vectors, the core mathematical routine in machine learning. The method can, in principle, be scaled to larger numbers of qubits, and may provide a new route to accelerate machine learning.

There are two main types of machine learning tasks [1], namely, supervised and unsupervised machine learning. In supervised machine learning, the learner is provided a set of training examples with features presented in the form of high-dimensional vectors and with corresponding labels to mark its category. The aim is to classify new examples based on these training sets. A simple example is a spam filter that sorts incoming Email into spam and nonspam messages by comparing the new Email with old Email already labeled by human. In unsupervised machine learning, the system aims to classify the data into different groups without prior information. An example of unsupervised machine learning is to recognize the object from a landscape background, i.e., to classify the pixels of the image into two groups—the object and the background. The core mathematical task for both the supervised and unsupervised machine learning algorithm is evaluating the distance and inner products between the high-dimensional vectors to analyze the similarity between vectors, which requires a time proportional to the size of the vectors on classical computers. With the rapidly growing data size in the modern world, such a task could pose a challenge even for the latest supercomputers.

Recently, it has been shown by Lloyd, Mohseni, and Rebentrost [2] that quantum computers, which are naturally good at manipulating vectors and matrices, could provide an asymptotically exponential speed-up over their classical counterparts in performing some machine learning tasks involving large vectors. Consider the task of assigning N-dimensional vectors to one of k clusters, each with M

representative samples; a quantum computer takes time $O[\log(MN)]$. The exponential speed-up of the quantum machine learning algorithm, and its potential wide applications, may make it one of the promising applications of quantum computers [2–4], in addition to Shor's factoring algorithm [5–9], quantum simulation [10–14], and the quantum algorithm for solving linear equation systems [15,16].

In this Letter, we report proof-of-principle demonstrations of the supervised and unsupervised quantum machine learning algorithm [2] on a small-scale photonic quantum processor. The core mathematical task is to assign two-, four-, and eight-dimensional vectors ($N = 2, 4, 8$) to two different clusters with one reference vector ($M = 1$) in each cluster. The two clusters are labeled as A and B, each with one reference sample vector $\vec{v}_A$ and $\vec{v}_B$, respectively. To classify the new sample which is represented by the vector $\vec{u}$, one common method is to calculate and compare the distance: $D_A = |\vec{u} - \vec{v}_A|$, and $D_B = |\vec{u} - \vec{v}_B|$. The new sample is assigned to the cluster to which the distance is smaller.

The vectors can be represented with quantum states with a normalization factor, i.e., $\vec{u} = |u||u\rangle$, $\vec{v} = |v||v\rangle$. To evaluate the distance $|\vec{u} - \vec{v}|$, a key step in the quantum machine learning algorithm [2] is to adjoin an ancillary qubit to the states of the reference and new vectors, creating an entangled state in the form

$$|\varphi\rangle = (|0\rangle_{\text{anc}}|u\rangle_{\text{new}} + |1\rangle_{\text{anc}}|v\rangle_{\text{ref}})/\sqrt{2}. \qquad (1)$$

Next, a single-qubit measurement is made on the ancillary qubit alone (the other qubits are simply ignored), projecting it onto the state

$$|\phi\rangle = (|u||0\rangle - |v||1\rangle)/\sqrt{|u|^2 + |v|^2}. \qquad (2)$$

The success probability p of this projective measurement can be estimated by repeated measurements. Remarkably, the inner product between $|u\rangle$ and $|v\rangle$ can be directly calculated from the p:

$$\langle u|v\rangle = (0.5 - p)(|u|^2 + |v|^2)/|u||v|, \qquad (3)$$

and the distance between $\vec{u}$ and $\vec{v}$ can then be obtained:

$$D = \sqrt{2p(|u|^2 + |v|^2)}. \qquad (4)$$

It is important to note that such an estimation can achieve a desired statistical accuracy simply by a sufficient number of repeated measurements, but is *independent* of the size (N) of the vectors, which gives a quantum speed-up.

This algorithm can be understood intuitively; the more difference between the pure states $|u\rangle$ and $|v\rangle$, the more entangled the Eq. (1) is. For examples, if $|u\rangle$ and $|v\rangle$ are identical, then the ancillary qubit is in the state $(|0\rangle + |1\rangle)/\sqrt{2}$, separable from the vector qubits, and $p = 0$, $D = 0$.

If $|u\rangle$ and $|v\rangle$ are orthogonal, then the Eq. (1) is maximally entangled, and $p = 0.5$, $D = \sqrt{|u|^2 + |v|^2}$.

In our experiment, we use single photons as qubits, where $|0\rangle$ and $|1\rangle$ are encoded with the photon's horizontal (H) and vertical (V) polarization, respectively. A schematic drawing of the experimental setup is illustrated in Fig. 1. Polarization-entangled photon pairs are generated by spontaneous parametric down-conversion [17] and prepared in the state

$$(|0\rangle_{\text{anc}}|0\rangle_{\text{vec}} + |1\rangle_{\text{anc}}|1\rangle_{\text{vec}})/\sqrt{2}. \qquad (5)$$

One photon (anc) is used as the ancillary qubit, and the other one (vec) will be used to encode the reference and incoming vectors using Sagnac-like interferometers (see Fig. 1).

To generate three- and four-photon entanglement resource states, we create two entangled photon pairs. Two single photons, one from each pair, are temporally and spatially superposed on a polarizing beam splitter (PBS). We select the events where one and only one single photon emits from each output. It can be concluded that the four photons are either all H polarized or V polarized, two cases that are quantum mechanically indistinguishable when all the other degrees of freedom of the photons are erased (see the caption of Fig. 1), thus projecting the four photons into the Greenberger-Horne-Zeilinger entangled state [18]:
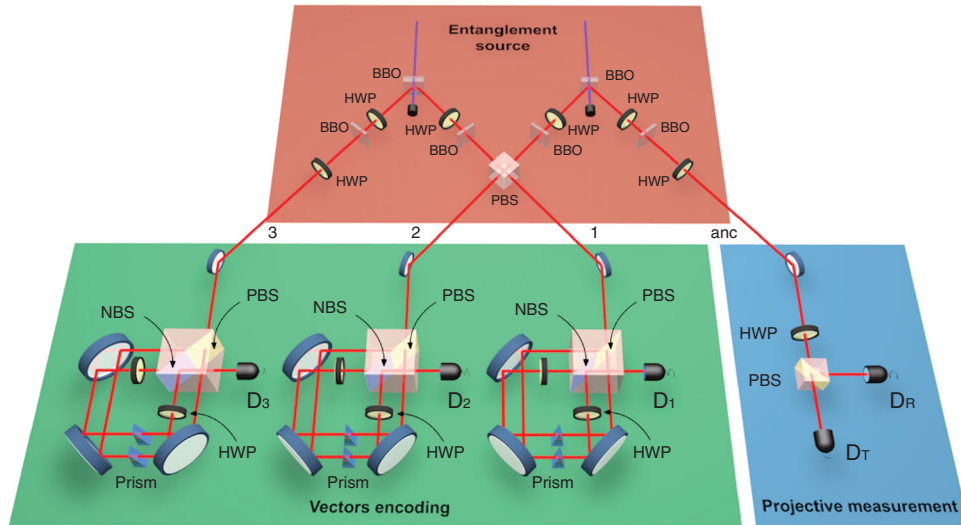

FIG. 1 (color). Experimental setup for quantum machine learning with photonic qubits. Ultraviolet laser pulses with a central wavelength of 394 nm, pulse duration of 120 fs, and a repetition rate of 76 MHz pass through two type-II β-barium borate (BBO) crystals with a thickness of 2 mm to produce two entangled photon pairs. The photons pass through pairs of birefringent compensators consisting of a 1-mm BBO crystal and a HWP to compensate the walk-off between horizontal and vertical polarization, and are prepared in the quantum state: $(|H\rangle|V\rangle + |V\rangle|H\rangle)/\sqrt{2}$. Two extra HWPs placed in arm 3 and anc are used to transform the state into $(|H\rangle|H\rangle + |V\rangle|V\rangle)/\sqrt{2}$. Two single photons, one from each pair, are temporally and spatially superposed on a PBS to generate a four-photon entangled state: $(|H\rangle|H\rangle|H\rangle|H\rangle + |V\rangle|V\rangle|V\rangle|V\rangle)/\sqrt{2}$. The photons 1, 2, and 3 are sent to Sagnac-like interferometers, where each single photon splits into two spatial modes by the PBS with regard to its polarization, and recombines on a nonpolarizing beam splitter (NBS). Various vectors are independently encoded into the two spatial modes using HWPs. The specially designed beam splitter cube is half-PBS coated and half-NBS coated. High-precision small-angle prisms are inserted for fine adjustments of the relative delay of the two different paths. The photons are detected by five single-photon detectors (quantum efficiency > 60%), and the two four-photon coincidence events, $D_3 D_2 D_1 D_T$ and $D_3 D_2 D_1 D_R$, are simultaneously registered by a homemade FPGA-based coincidence unit.

$$(|0\rangle_{\text{anc}}|000\rangle_{\text{vec}} + |1\rangle_{\text{anc}}|111\rangle_{\text{vec}})/\sqrt{2}. \quad (6)$$

By projecting one of the four photons into $(|H\rangle + |V\rangle)/\sqrt{2}$, we can reduce the four-photon state (6) to a three-photon entangled state:

$$(|0\rangle_{\text{anc}}|00\rangle_{\text{vec}} + |1\rangle_{\text{anc}}|11\rangle_{\text{vec}})/\sqrt{2}. \quad (7)$$

The two-, three-, and four-photon entangled states, Eqs. (5)–(7), are the entanglement resource used for the classification of the two-, four-, and eight-dimensional vectors, respectively. We characterize the created multiphoton entangled state using the method of entanglement witness. We obtain the fidelity [19] for the two-, three-, and

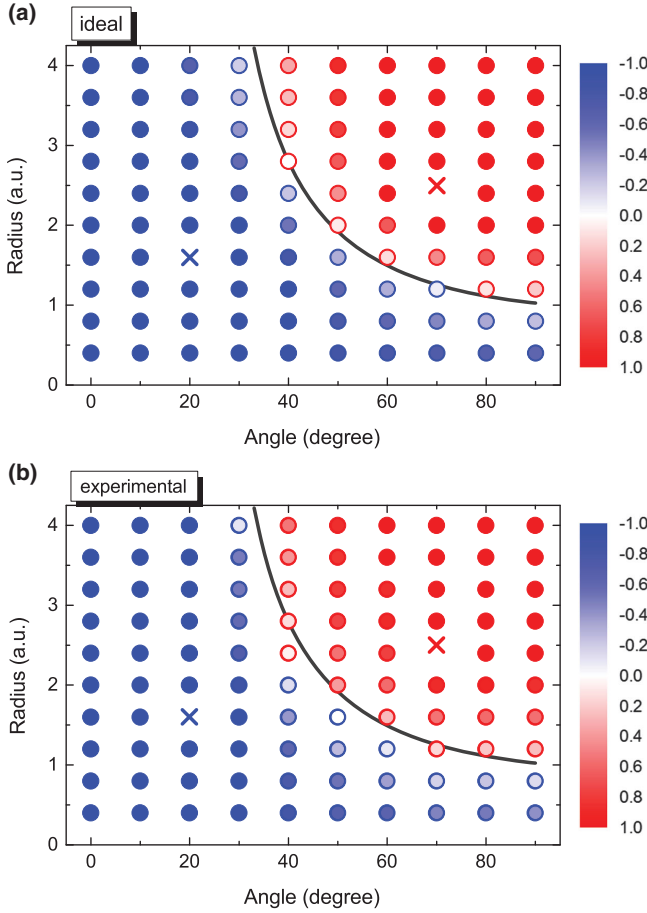

(a) ideal

(b) experimental

FIG. 2 (color). Theoretical prediction (a) and experimental results (b) for classifying two-dimensional vectors into two clusters. The red and blue cross are reference vectors. The evaluated value of difference of the distances of each tested vector to the two reference vectors is coded as the fill color. The result of classification is coded as the edge color (blue = A, red = B). The gray line is where the distances are the same in theory. The data acquisition time is 1 sec for each vector, collecting about 10000 events. The statistical standard deviation is much smaller than the error caused by the imperfection of the entanglement state; thus, error bars are omitted. The data are represented in polar coordinate.

four-photon entangled states to be 0.94, 0.73, and 0.75, respectively, thus, prove the presence of genuine multipartite entanglement [20].

A 2^n-dimensional vector is encoded with the polarization state of n photonic qubits. For example, a four-dimensional vector, (3.42, 1.24, 1.97, 0.72), is represented by the composite quantum state of two single photons with normalization,

$$|u_2 u_1| \times |u_2 u_1\rangle = 4.2 \times (0.866|0\rangle + 0.5|1\rangle)$$
$$\otimes (0.94|0\rangle + 0.342|1\rangle). \quad (8)$$

To encode these vectors into the entanglement resource states, Eqs. (5)–(7), we send the single photons through a PBS where the photon is split into two spatial modes according to its polarization. At the two separate spatial modes, controlled unitary operations can be implemented deterministically and independently [21]. Thus, we can transform, for instance, the two-photon entangled state (5) into $(|0\rangle|u_1\rangle_{\text{new}} + |1\rangle|v_1\rangle_{\text{ref}})/\sqrt{2}$, where the state $|u_1\rangle$ and $|v_1\rangle$ can be arbitrarily set using wave plates. The two spatial modes are then recombined on a nonpolarizing beam splitter. In this way, we create the following two-, three-, and four-photon entangled states in the form of Eq. (1):

$$|\varphi_2\rangle = (|0\rangle_{\text{anc}}|u_1\rangle_{\text{new}} + |1\rangle_{\text{anc}}|v_1\rangle_{\text{ref}})/\sqrt{2}$$
$$|\varphi_4\rangle = (|0\rangle_{\text{anc}}|u_2 u_1\rangle_{\text{new}} + |1\rangle_{\text{anc}}|v_2 v_1\rangle_{\text{ref}})/\sqrt{2}$$
$$|\varphi_8\rangle = (|0\rangle_{\text{anc}}|u_3 u_2 u_1\rangle_{\text{new}} + |1\rangle_{\text{anc}}|v_3 v_2 v_1\rangle_{\text{ref}})/\sqrt{2} \quad (9)$$

for classifying two-, four-, and eight-dimensional vectors, respectively.

TABLE I. Experimental results for classifying four-dimensional vectors into two clusters. Reference vector A is $(1, 0, 0, 0)$ and B is $(0, 0, 1, 1)$. The data acquisition time is 2 min for each vector, collecting about 500 events.

	Test vectors	$\frac{D_A - D_B}{\text{Theory}}$	Exp.	Group	Correct?
1	(2.00, 0.00, 0.00, 0.00)	−1.45	−0.93	A	✓
2	(0.00, 0.00, 0.00, 2.00)	0.82	0.50	B	✓
3	(0.35, 0.20, 0.00, 0.00)	−0.79	−0.71	A	✓
4	(0.23, 0.19, 0.08, 0.07)	−0.54	−0.51	A	✓
5	(1.32, 3.62, 1.57, 4.32)	0.74	0.48	B	✓
6	(0.15, 0.17, 0.82, 0.98)	1.26	0.72	B	✓
7	(0.18, 0.10, 1.02, 0.59)	0.98	0.76	B	✓
8	(0.97, 0.17, 0.17, 0.03)	−1.37	−0.93	A	✓
9	(0.68, 0.25, 0.00, 0.00)	−1.18	−0.79	A	✓
10	(0.83, 0.48, 1.44, 0.83)	0.67	0.17	B	✓
11	(1.27, 1.06, 3.48, 2.92)	1.13	0.76	B	✓
12	(0.40, 0.40, 0.40, 0.40)	−0.10	−0.26	A	✓
13	(0.09, 0.15, 0.49, 0.85)	0.80	0.55	B	✓
14	(0.10, 0.55, 0.06, 0.32)	−0.19	−0.28	A	✓
15	(1.94, 0.34, 0.34, 0.06)	−1.22	−1.10	A	✓
16	(3.42, 1.24, 1.97, 0.72)	−0.34	−0.39	A	✓
17	(0.66, 0.00, 1.80, 0.00)	0.40	−0.02	A	✗

TABLE II.   Experimental results for classifying eight-dimensional vectors into two clusters. Reference vector A is $(1, 0, 0, 0, 0, 0, 0, 0)$ and B is $(0, 0, 0, 0, 0, 0, 0, 1)$. The data acquisition time is 4 min for each vector, collecting about 500 events.

Test vectors		$\frac{D_A - D_B}{\text{Theory}}$ Exp.		Group	Correct?
1	(2.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00)	−1.24	−0.84	A	✓
2	(0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.60)	0.77	0.55	B	✓
3	(1.77, 0.00, 0.00, 0.00, 1.24, 0.00, 0.00, 0.00)	−0.92	−0.52	A	✓
4	(0.40, 0.23, 0.11, 0.06, 0.03, 0.02, 0.01, 0.01)	−0.45	−0.14	A	✓
5	(0.00, 0.00, 1.23, 1.23, 0.00, 0.00, 0.33, 0.33)	0.17	0.10	B	✓
6	(0.30, 0.03, 0.30, 0.03, 1.12, 0.10, 1.12, 0.10)	−0.11	−0.24	A	✓
7	(0.42, 0.90, 0.35, 0.76, 0.00, 0.00, 0.00, 0.00)	−0.28	−0.21	A	✓
8	(0.54, 0.54, 0.00, 0.00, 0.54, 0.54, 0.00, 0.00)	−0.43	−0.50	A	✓
9	(0.11, 1.24, 0.19, 2.15, 0.06, 0.72, 0.11, 1.24)	0.40	−0.17	A	✗

Figures 2(a) and 2(b) display theoretical prediction (ideal) and experimental results of entanglement-based classification of two-dimensional vectors. The randomly chosen reference vectors are $\vec{v}_A = (1.50, 0.55)$ and $\vec{v}_B = (0.86, 2.35)$, plotted in polar coordinate as blue and red rectangular crosses, respectively. For each new vector $|u_i\rangle_{\text{new}}$, $i = 1, 2, \ldots, 100$, two-photon entangled states $|\varphi_2\rangle$ are constructed, and the distances from $\vec{u}_i$ to $\vec{v}_A$ and $\vec{v}_B$ (denoted by D_A and D_B) are evaluated from the success probability of the projective measurements on the ancillary photon.

The difference of the distances, $D_A - D_B$, is color coded in the fill color of each of the data points in Figs. 2(a) and 2(b). The sign of $D_A - D_B$ dictates the result of the classification: if $D_A - D_B < 0$, it is categorized to cluster A (plotted as the blue edge color); if $D_A - D_B > 0$, it is categorized to cluster B (plotted as red edge color). The boundary of the two clusters (where $D_A = D_B$) is illustrated as the gray line in Figs. 2(a) and 2(b). It can be seen that, of the 100 tested

samples, two are experimentally misclassified. The misclassification happens for vectors close to the boundary where the absolute error (with an average of ∼0.27), caused by the imperfect two-photon entanglement and dark counts of the single-photon detectors, becomes comparable to $|D_A - D_B|$.

Similar methods can be applied to the classifications of four- and eight-dimensional vectors based on the construction of three- and four-photon entanglement, with the experimental results listed in Tables I and II, respectively. The precision of distance evaluation is affected by the state fidelity (∼75%) of the multiphoton entangled state, which is lower compared to that of the two-photon entangled state (∼94%), mainly caused by double pair emission in parametric down-conversion, the imperfect interference of independent photons on the PBS [18], and the phase fluctuations in the Sagnac interferometers. Among the randomly selected 17 and 9 vectors listed in Tables I and II, respectively, there is one sample misclassified.
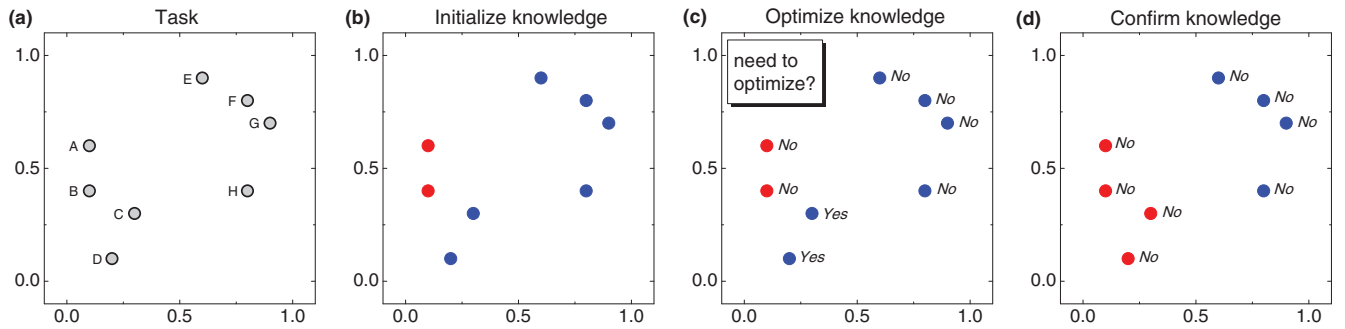

FIG. 3 (color).   Demonstration of unsupervised machine learning. (a) Eight gray circles, labeled from A to H, are two-dimensional vectors to be classified. (b) A random classification is initialized, where A and B belong to the red group and C, D, E, F, G, H belong to blue group. (c) We use the entanglement-based method presented in this Letter to experimentally evaluate the distance from each vector to the other vectors within a group. Then the mean distance to both the red and the blue group is calculated, $D_{A\text{-red}} = 0.12$, $D_{A\text{-blue}} = 0.67$, $D_{B\text{-red}} = 0.12$, $D_{B\text{-blue}} = 0.65$, $D_{C\text{-red}} = 0.41$, $D_{C\text{-blue}} = 0.68$, $D_{D\text{-red}} = 0.33$, $D_{D\text{-blue}} = 0.58$, $D_{E\text{-red}} = 0.73$, $D_{E\text{-blue}} = 0.59$, $D_{F\text{-red}} = 0.84$, $D_{F\text{-blue}} = 0.57$, $D_{G\text{-red}} = 0.91$, $D_{G\text{-blue}} = 0.57$, $D_{H\text{-red}} = 0.77$, $D_{H\text{-blue}} = 0.50$. It can be seen that C and D are closer to the red group but were wrongly classified into the blue group, whose labels are therefore changed. The labels of the other six vectors remain unchanged. (d) The optimizing process is repeated in the new configuration until there is no change. In this configuration every vector is in the group with a closer mean distance, and the system can confirm the configuration as an ultimate classification result.

The quantum mechanical way of evaluating the vector distance demonstrated above is the core mathematical subroutine for other machine learning tasks, for example, the supervised nearest-neighbor algorithm and unsupervised machine learning algorithm. In the supervised nearest-neighbor algorithm, each test vector is analyzed by evaluating the distance between itself and all the training vectors, and then categorized into the group of the nearest training vector. When new training vectors are offered, the system will adjust the judgment of classification by analyzing the distances in the new configuration. An example with training sample $M = 2$ is shown in Supplemental Fig. S1 [22].

In unsupervised machine learning, no training vectors are provided, and the system needs to realize a reasonable classification by iterating to calculate the distance between different vectors. The algorithm includes three steps. (i) Initialize a random classification. (ii) For each vector v_i, the learner calculates the distance between v_i and all vectors in a group. The v_i is classified into a group to which the average distance is minimal. (iii) Repeat step 2 until no vector needs to change its group. An example with $M = 4$ is demonstrated in Fig. 3.

Note that the current experimental scheme can, in principle, achieve an exponential speed-up with respect to the dimension N of the vectors, but not to the number of training samples M. To demonstrate a speed-up in numbers of manifold vectors M, future studies are planned to design quantum circuits involving $M + 1$ level qudits. High-dimensional quantum states can be encoded using, for example, a photons' degree of freedom of orbital angular momentum [23].

In summary, we have performed the first experimental demonstration of machine learning on a photonic quantum computer. Our work demonstrates that the manipulation of high-dimensional vectors and the estimation of the distance and inner product between vectors, a ubiquitous task in machine learning, can be naturally done with quantum computers, thus proving the suitability and potential power of quantum machine learning. The ability of manipulating large vectors—combined with previously realized methods for solving systems of linear equations [15,16] and Hamiltonian simulation [24]—on quantum computers, may provide a useful quantum toolkit for dealing with the "big data."

*cylu@ustc.edu.cn
†eidos@ustc.edu.cn
‡nlliu@ustc.edu.cn

[1] M. Mohri, A. Rostamizadeh, and A. Talwalkar, *Foundations of Machine Learning* (MIT Press, Cambridge, MA, 2012).

[2] S. Lloyd, M. Mohseni, and P. Rebentrost, arXiv:1307.0411.

[3] M. Sasaki and A. Carlini, Phys. Rev. A **66**, 022303 (2002).

[4] E. Aimeur, G. Brassard, and S. Gambs, Mach. Learn. **90**, 261 (2013).

[5] P. Shor, SIAM J. Comput. **26**, 1484 (1997).

[6] L. M. K. Vandersypen, M. Steffen, G. Breyta, C. S. Yannoni, M. H. Sherwood, and I. L. Chuang, Nature (London) **414**, 883 (2001).

[7] C.-Y. Lu, D. E. Browne, T. Yang, and J.-W. Pan, Phys. Rev. Lett. **99**, 250504 (2007).

[8] B. P. Lanyon, T. J. Weinhold, N. K. Langford, M. Barbieri, D. F. V. James, A. Gilchrist, and A. G. White, Phys. Rev. Lett. **99**, 250505 (2007).

[9] E. Lucero *et al.*, Nat. Phys. **8**, 719 (2012).

[10] S. Lloyd, Science **273**, 1073 (1996).

[11] I. Bloch, J. Dalibard, and S. Nascimbéne, Nat. Phys. **8**, 267 (2012).

[12] R. Blatt and C. F. Roos, Nat. Phys. **8**, 277 (2012).

[13] A. Aspuru-Guzik and P. Walther, Nat. Phys. **8**, 285 (2012); C.-Y. Lu, W.-B. Gao, O. Gühne, X.-Q. Zhou, Z.-B. Chen, and J.-W. Pan, Phys. Rev. Lett. **102**, 030502 (2009); B. P. Lanyon *et al.*, Nat. Chem. **2**, 106 (2010); X.-S. Ma, B. Dakic, W. Naylor, A. Zeilinger, and P. Walther, Nat. Phys. **7**, 399 (2011); L. Sansoni, F. Sciarrino, G. Vallone, P. Mataloni, A. Crespi, R. Ramponi, and R. Osellame, Phys. Rev. Lett. **108**, 010502 (2012).

[14] A. A. Houck, H. E. Türeci, and J. Koch, Nat. Phys. **8**, 292 (2012).

[15] A. W. Harrow, A. Hassidim, and S. Lloyd, Phys. Rev. Lett. **103**, 150502 (2009).

[16] X.-D. Cai, C. Weedbrook, Z.-E. Su, M.-C. Chen, M. Gu, M.-J. Zhu, L. Li, N.-L. Liu, C.-Y. Lu, and J.-W. Pan, Phys. Rev. Lett. **110**, 230501 (2013); J. Pan, Y.-D. Cao, X.-W. Yao, Z.-K. Li, C.-Y. Ju, H.-W. Chen, X.-H. Peng, S. Kais, and J.-F. Du, Phys. Rev. A **89**, 022313 (2014); S. Barz, I. Kassal, M. Ringbauer, Y. O. Lipp, B. Dakić, A. Aspuru-Guzik, and P. Walther, Sci. Rep. **4**, 6115 (2014).

[17] P. G. Kwiat, K. Mattle, H. Weinfurter, A. Zeilinger, A. V. Sergienko, and Y. Shih, Phys. Rev. Lett. **75**, 4337 (1995).

[18] J.-W. Pan, Z.-B. Chen, C.-Y. Lu, H. Weinfurter, A. Zeilinger, and M. Żukowski, Rev. Mod. Phys. **84**, 777 (2012).

[19] The state fidelity is a measure of to what extent the desired state is generated. It can be calculated by the overlap of the experimentally produced state with the ideal one.

[20] O. Gühne, C.-Y. Lu, W.-B. Gao, and J.-W. Pan, Phys. Rev. A **76**, 030305 (2007).

[21] X.-Q. Zhou, P. Kalasuwan, T. C. Ralph, and J. L. O'Brien, Nat. Photonics **7**, 223 (2013).

[22] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.114.110504 for data of supervised nearest-neighbour machine learning algorithm.

[23] R. Fickler, R. Lapkiewicz, W. N. Plick, M. Krenn, C. Schaeff, S. Ramelow, and A. Zeilinger Science **338**, 640 (2012); X.-L. Wang, X.-D. Cai, Z.-E. Su, M.-C. Chen, D. Wu, L. Li, N.-L. Liu, C.-Y. Lu, and J.-W. Pan, Nature (London) **518**, 516 (2015).

[24] D. W. Berry, G. Ahokas, R. Cleve, and B. C. Sanders, Commun. Math. Phys. **270**, 359 (2007).