Quantum-Enhanced Data Classification with a Variational Entangled Sensor Network

Yi Xia,[1] Wei Li,[2] Quntao Zhuang,[3,1] and Zheshen Zhang[2,1,*]

[1]*James C. Wyant College of Optical Sciences, University of Arizona, Tucson, Arizona 85721, USA*
[2]*Department of Materials Science and Engineering, University of Arizona, Tucson, Arizona 85721, USA*
[3]*Department of Electrical and Computer Engineering, University of Arizona, Tucson, Arizona 85721, USA*

Variational quantum circuits (VQCs) built upon noisy intermediate-scale quantum (NISQ) hardware, in conjunction with classical processing, constitute a promising architecture for quantum simulations, classical optimization, and machine learning. However, the required VQC depth to demonstrate a quantum advantage over classical schemes is beyond the reach of available NISQ devices. Supervised learning assisted by an entangled sensor network (SLAEN) is a distinct paradigm that harnesses VQCs trained by classical machine-learning algorithms to tailor multipartite entanglement shared by sensors for solving practically useful data-processing problems. Here, we report the first experimental demonstration of SLAEN and show an entanglement-enabled reduction in the error probability for classification of multidimensional radio-frequency signals. Our work paves a new route for quantum-enhanced data processing and its applications in the NISQ era.

Subject Areas: Optics, Quantum Information

I. INTRODUCTION

The convergence of quantum information science and machine learning (ML) has endowed radically new capabilities for solving complex physical and data-processing problems [1–9]. Many existing quantum ML schemes hinge on large-scale fault-tolerant quantum circuits composed of, e.g., quantum random access memories. At present, however, the available noisy intermediate-scale quantum (NISQ) devices [10,11] hinder these quantum ML schemes to achieve an advantage over classical ML schemes. Recent developments in hybrid systems [9,12] comprising classical processing and variational quantum circuits (VQCs) open an alternative avenue for quantum ML. In this regard, a variety of hybrid schemes have been proposed, including quantum approximate optimization [13], variational quantum eigensolvers [14], quantum multiparameter estimation [15], and quantum kernel estimators and variational quantum models [4,5]. On the experimental front, hybrid schemes have been implemented to seek the ground state of quantum systems [14,16], to perform data classification [4], to unsample a quantum circuit [17], and to solve the MAXCUT problem [18,19]. The finite quantum coherence time and circuit depths of

state-of-the-art NISQ platforms, however, hold back a near-term quantum advantage over classical ML schemes. An imperative objective for quantum ML is to harness NISQ hardware to benefit practically useful applications [2].

II. SUPERVISED LEARNING ASSISTED BY AN ENTANGLED SENSOR NETWORK

A multitude of data-processing scenarios, such as classification of images captured by cameras [20], target detection through a phased array [21], and identification of molecules [22], encompass sensors for data acquisition. Recent theoretical [23–29] and experimental [30–32] advances in distributed quantum sensing have unleashed the potential for a network of entangled sensors to outperform classical separable sensors in capturing global features of an interrogated object. Such a capability endowed by distributed quantum sensing creates an opportunity to further utilize VQCs to configure the entangled probe state shared by the sensors to enable a quantum advantage in data-processing problems.

Supervised learning assisted by an entangled sensor network (SLAEN) [33] is such a hybrid quantum-classical framework empowered by entangled sensors configured by a classical support-vector machine (SVM) for quantum-enhanced high-dimensional data classification, as sketched in Fig. 1(a). SLAEN employs a VQC parametrized by $\boldsymbol{v}$ to create an entangled probe state $\hat{\rho}_E$ shared by M quantum sensors. The sensing attempt at the mth sensor is modeled by running the probe state through a quantum channel, $\Phi(\alpha_m)$, where the information about the object is embedded
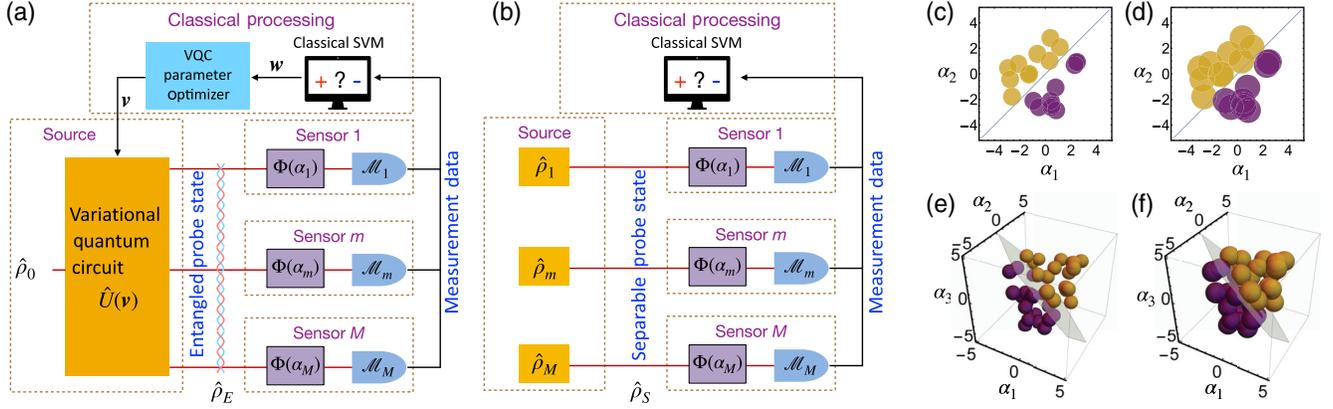
FIG. 1.   Schematics of SLAEN and classical classifier with sample data sets. (a) In SLAEN, a VQC is configured to generate an entangled probe state. In classical processing, measurement data are utilized to train a classical SVM, whose hyperplane w is mapped to the VQC setting v by the VQC parameter optimizer. (b) Classical classifier using only a classical SVM. (c,d) 2D data acquired by two sensors, applicable to rf-field direction classification. (e,f) 3D data acquired by three sensors, applicable to rf-field mean-amplitude classification. Circle or sphere: data point with radius representing standard deviation of estimation uncertainty. (c,e) Entangled sensors with a clear error-probability reduction over (d,f) classical separable sensors.

in the parameter α_m. A measurement $\mathcal{M}_m$ on the output quantum state from the channel then yields $\tilde{\alpha}_m$ as the measurement data. To label the interrogated object, a classical SVM chooses a hyperplane parametrized by w to separate the measurement data into two classes in an M-dimensional space. To learn the optimum hyperplane and the configuration of the VQC that produces the optimum entangled probe state under a given classification task, the sensors first probe training objects with known labels, and the measurement data and the true labels are used to optimize the hyperplane w of the SVM. Then, the VQC parameter optimizer maps $w \to v$, which in turn configures the VQC to generate an entangled probe state $\hat{\rho}_E = \hat{U}(v)\hat{\rho}_0\hat{U}^\dagger(v)$ that minimizes the measurement noise subject to the chosen hyperplane. As a comparison, Fig. 1(b) sketches a conventional classical classifier that solely relies on a classical SVM trained by the measurement data obtained by separable sensors to seek the optimum hyperplane for classification. By virtue of the entanglement-enabled noise reduction, SLAEN yields a substantially lower error probability than that achieved by the classical classifier, which is illustrated and compared in Figs. 1(c)–1(f) for two classification problems in, respectively, a 2D data space and a 3D data space.

III. EXPERIMENT

We demonstrate SLAEN in a quantum optics platform based on continuous-variable (CV) entanglement. The experiment endeavors to classify a feature embedded in a set of rf signals: $\{\mathcal{E}_m(t) = E_m \cos(\omega_c t + \varphi_m)\}_{m=1}^M$, where $E \equiv \{E_m\}_{m=1}^M$ and $\varphi \equiv \{\varphi_m\}_{m=1}^M$ are, respectively, the rf amplitudes and phases at the $M = 3$ sensors, and ω_c is the rf carrier frequency. The class label y is determined by a joint function of amplitudes and phases: $y = F(E, \varphi)$.

The experimental setup is sketched in Fig. 2. An optical parametric amplifier source emits a single-mode squeezed state represented by the annihilation operator $\hat{b}$. To acquire data, a VQC prepares an entangled probe state, described by $\{\hat{b}_m\}_{m=1}^M$, by applying a unitary operation $\hat{U}(v)$ on $\hat{b}$. The VQC setting is entailed in $v \equiv \{v_m, \phi_m\}_{m=1}^M$, where v_m is the power ratio of the squeezed state sent to the mth sensor, satisfying $\sum_{m=1}^M v_m = 1$, and ϕ_m is a phase shift imparted on the quantum state at the mth sensor. The VQC is composed of two variable beam splitters (VBSs) and three phase shifters. A VBS comprises two half-wave plates (H), a quarter-wave plate (Q), a phase modulator (PM), and a polarizing beam splitter (PBS). The PM controls the splitting ratio of the VBS and thus determines v_m, while the phase shift ϕ_m is controlled by a rf signal delay (see the Appendix B for details). At the mth sensor, an electro-optic modulator (EOM) converts the rf signal into a displacement $\alpha_m \propto E_m \sin \varphi_m$ on the phase quadrature $\hat{p}_m \equiv (\hat{b}_m - \hat{b}_m^\dagger)/2i$. Three homodyne detectors then measure the quadrature displacements, and the measurement data are diverted to a classical processing unit for training, classification, and VQC setting optimization.

SLAEN consists of a training stage and a utilization stage. The training stage is aimed at using N training data points $\{E^{(n)}, \varphi^{(n)}, y^{(n)}\}_{n=1}^N$ supplied to the sensors to optimize the hyperplane used by the SVM and the entangled probe state. Note that $y^{(n)} \in \{-1, +1\}$ is the true label for the nth training data point. The training data point leads to the homodyne measurement data $\tilde{\alpha}^{(n)}$ from the sensors. Here, $\tilde{\alpha}^{(n)}$ and $y^{(n)} \in \{-1, +1\}$ are the only information available to the classical processing unit. For a hyperplane specified by $\{w \equiv \{w_m\}_{m=1}^M, b\}$, we define a cost function
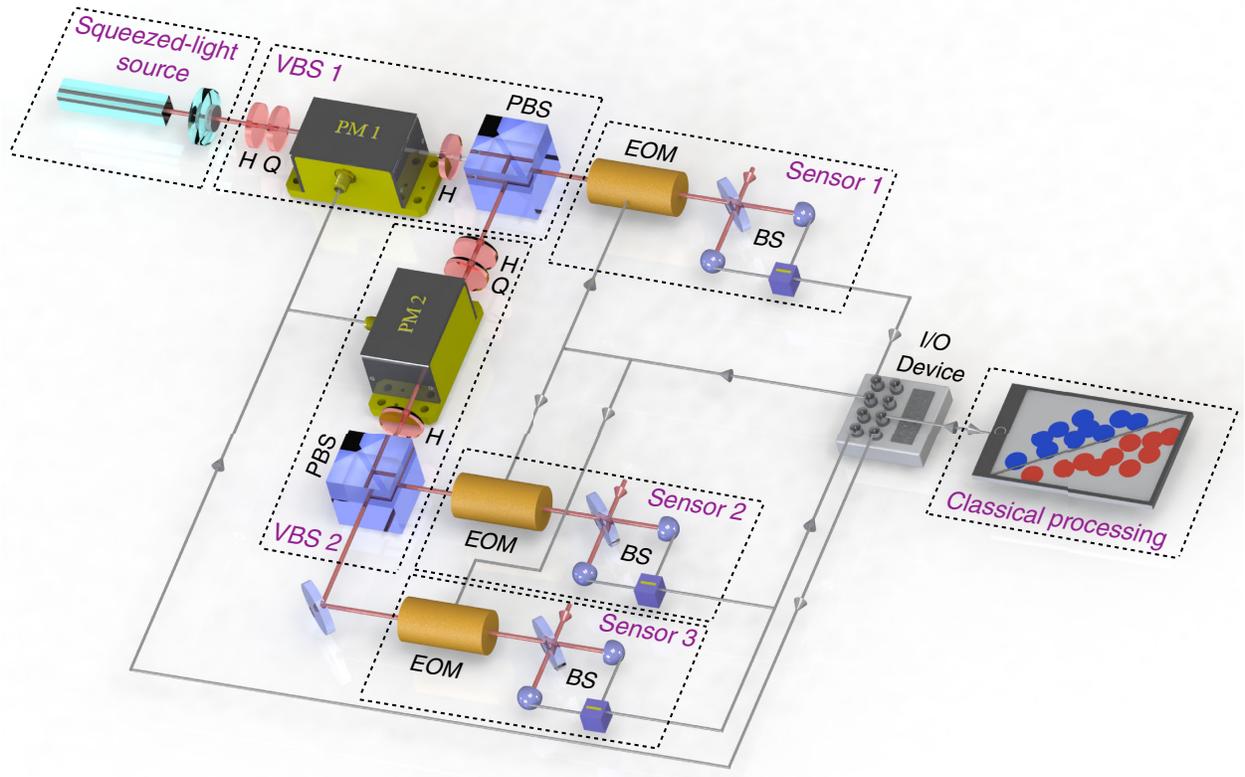
FIG. 2.   Experimental diagram. Squeezed light processed by two VBSs, each composed of two H plates, a Q plate, a PM, and a PBS, generating a three-partite entangled probe state. Each sensor comprises an EOM and a balanced homodyne measurement setup. Measurement data were acquired by an I/O device and processed on a classical computer for training and data classification. During training, classical processing controls VBSs and EOMs through the I/O device.

$$\mathcal{E}_{\lambda}(\boldsymbol{w}, b) = \sum_{n=1}^{N} |1 - y^{(n)}(\boldsymbol{w} \cdot \tilde{\boldsymbol{\alpha}}^{(n)} + b)|_{+} + \lambda \|\boldsymbol{w}\|^2, \quad (1)$$

where $|x|_{+}$ equals x for $x \geq 0$ and zero otherwise, $\| \cdot \|$ is the usual two-norm, and $\lambda\|\boldsymbol{w}\|^2$ is used to avoid overfitting. The $\boldsymbol{w} \cdot \tilde{\boldsymbol{\alpha}}^{(n)}$ term represents a weighted average over the measurement data acquired by different sensors. It is the weighted average that benefits from using multipartite entanglement to reduce the measurement noise [31,33]. Only the support vectors, i.e., points close to the hyperplane with $y^{(n)}(\boldsymbol{w} \cdot \tilde{\boldsymbol{\alpha}}^{(n)} + b) \leq 1$, contribute nontrivially to the cost function. The rationale behind constructing such a cost function is that errors primarily occur on support vectors in a classification task; thus, accounting for the deviations of all data points from the hyperplane in the cost function is nonideal.

To enable efficient minimization of the cost function, we adopt a stochastic optimization approach in which the hyperplane and the VQC setting are updated in each training step consuming a single data point. Suppose the optimized hyperplane is $\{\boldsymbol{w}^{(n-1)}, b^{(n-1)}\}$ after $(n-1)$ training steps. Prior to updating the hyperplane in the nth training step, the inferred label is derived by

$\tilde{y}^{(n)} = \text{sign}(\boldsymbol{w}^{(n-1)} \cdot \tilde{\boldsymbol{\alpha}}^{(n)} + b^{(n-1)})$. Using a simultaneous perturbation stochastic approximation (SPSA) algorithm, the hyperplane is updated to $\{\boldsymbol{w}^{(n)}, b^{(n)}\}$ (see Appendix A for algorithm details). Once an updated hyperplane is found, the VQC optimizer performs the mapping $\boldsymbol{w}^{(n)} \rightarrow \boldsymbol{v}^{(n)}$ to configure the VQC so that its generated entangled probe state minimizes the measurement noise subject to the current hyperplane. Specifically, one desires that the virtual mode $\hat{b}_v \equiv \sum_{m=1}^{M} w_m^{(n)} \hat{b}_m$, whose phase-quadrature measurement outcome constitutes the $\boldsymbol{w}^{(n)} \cdot \tilde{\boldsymbol{\alpha}}^{(n+1)}$ term in $\tilde{y}^{(n+1)}$, is identical to the original squeezed-light mode $\hat{b}$ so that the overall uncertainty in labeling is minimized. This goal is accomplished by setting $\sqrt{v_m^{(n)}} \exp{(i\phi_m^{(n)})} \propto w_m^{(n)}$ in the VQC parameter optimizer. Physically, this is the noise-reduction mechanism, stemming from the quantum correlations between the measurement noise at different sensors, that gives rise to SLAEN's quantum advantage over the classical classifier in which the measurement noise at different sensors is independently subject to the standard quantum limit. After N training steps, the cost function is near its minimum with the hyperplane $\{\boldsymbol{w}^{\star}, b^{\star}\} \equiv \{\boldsymbol{w}^{(N)}, b^{(N)}\}$, and the VQC setting $\boldsymbol{v}^{\star} \equiv \boldsymbol{v}^{(N)}$. Then, in the utilization stage, SLAEN configures the VQC using $\boldsymbol{v}^{\star}$ and classifies the

measurement data $\tilde{\alpha}$ with an unknown label using the optimized hyperplane $w^{\star}$:

$$\tilde{y} = \text{sign}(w^{\star} \cdot \tilde{\alpha} + b^{\star}). \tag{2}$$

SLAEN is a versatile framework capable of tailoring the entangled probe state and the classical SVM to enhance the performance of multidimensional data-classification tasks. In our experiment, SLAEN first copes with 2D data acquired by two entangled sensors, as illustrated in Fig. 1(c). As an example and useful application for 2D data classification, we demonstrate the classification of the incident direction of an emulated rf field. To this end, the training commences with an initial hyperplane specified by $\{w^{(0)} = (\sqrt{0.50}, \sqrt{0.50}), b^{(0)} = 0.70\}$, which is mapped to an initial VQC setting $v^{(0)} = \{v_1 = 0.5, v_2 = 0.5, v_3 = 0, \phi_1 = 0, \phi_2 = 0, \phi_3 = 0\}$. The training stage comprises 200 steps each, using a training data point with randomly generated rf-field phases and an associated label $\{\varphi^{(n)}, y^{(n)}\}_{n=1}^{200}$, while the rf-field amplitudes are fixed equal at all sensors. Applying the training data $\varphi^{(n)}$ on the EOMs at the two sensors leads to quadrature displacements $\alpha^{(n)} = \{\alpha_1^{(n)}, \alpha_2^{(n)}\}$, whose components are each chosen to follow a uniform distribution in $[-4, 4]$ (in shot-noise units). The signal-to-noise ratio of the data set is tuned by excluding the data points within a margin of ϵ from the hyperplane, while the total number of training data points is fixed at 200. In doing so, the signal-to-noise ratio increases as ϵ increases. The true labels for the rf-field directions are derived by the rf-phase gradient: $y^{(n)} = \text{sign}(\varphi_1^{(n)} - \varphi_2^{(n)}) = \text{sign}(w_t \cdot \alpha^{(n)})$, where $\{w_t = (\sqrt{1/2}, -\sqrt{1/2}), b_t = 0\}$ parametrize the true hyperplane. The true labels are disclosed while $\{w_t, b_t\}$ and $\alpha^{(n)}$ are kept unknown to SLAEN. The optimization for the SVM hyperplane and the VQC setting then follows.

As a performance benchmark, we train the classical classifier, using the identical training data in training SLAEN, to undertake the 2D data-classification task. Unlike SLAEN, the classical classifier uses a separable probe state $\hat{\rho}_S$ to acquire the measurement data, which are then used to train the classical SVM to seek a hyperplane that minimizes the classification error probability. In the experiment, the squeezed-light source is turned off while applying the same training data as those used for SLAEN, thereby ensuring an equitable performance comparison. The initial hyperplane prior to the training is randomly picked as $\{w^{(0)} = (0.67, 0.74), b^{(0)} = 0.39\}$. In the absence of entanglement-enabled noise reduction, a higher error probability is anticipated for the classical classifier, as illustrated in Fig. 1(d).

The effectiveness of the training for SLAEN and the classical classifier is demonstrated by the converging error probabilities measured at different training steps, as plotted in Fig. 3(a). The inset describes the VQC parameters being

optimized. The convergence of the error probabilities beyond 100 training steps indicates that near-optimum settings for the hyperplanes and the VQC have been found. With such optimized parameters, SLAEN is able to generate an entangled probe state that minimizes the measurement noise, as illustrated in Fig. 1(c) and compared to Fig. 1(d) for the case of the classical classifier by a set of sample data points represented by the circles, whose radii correspond to the standard deviation of estimation uncertainty.

SLAEN and the classical classifier are next trained to tackle 3D data-classification problems. As an example, we demonstrate the classification of the sign for the rf-field mean amplitude across three sensors. The training in either scenario uses 390 data points $\{E^{(n)}, y^{(n)}\}_{n=1}^{390}$ with randomly generated rf-field amplitudes, while the rf phases are fixed at $\varphi^{(n)} = 0$. The true labels are then given by $y^{(n)} = \text{sign}(\sum_{m=1}^{3} E_m^{(n)}) = \text{sign}(w_t \cdot \alpha^{(n)})$, where $\{w_t = (\sqrt{1/3}, \sqrt{1/3}, \sqrt{1/3}), b_t = 0\}$ specify the true hyperplane, which is unknown to SLAEN and the classical classifier. The error probabilities during training for both scenarios are plotted in Fig. 3(d), with its inset describing the VQC parameters being optimized. The error probabilities converge after 250 training steps, indicating that near-optimum settings for the hyperplanes and the VQC have been found. Once both are trained, SLAEN shows a clear error-probability advantage over that of the classical classifier, as observed in Fig. 3(d) and intuitively illustrated in Figs. 1(e) and 1(f).

The trajectories of the evolving hyperplane $\{w^{(n)}, b^{(n)}\}$ during training are plotted in Figs. 3(b) and 3(c) for 2D data classification and in Figs. 3(e) and 3(f) for 3D data classification. The hexagrams entail the optimum hyperplane parameters. The hyperplane parameters approach the optimum with a decreasing error probability during training, as anticipated. Notably, the optimized hyperplanes obtained by SLAEN are considerably closer to the true hyperplanes, i.e., the optimum solutions, than those attained by the classical classifier, thanks to SLAEN's reduced measurement noise. To further investigate SLAEN's improved accuracy to problem solutions, we randomly generate 50 sets of initial hyperplanes for SLAEN and the classical classifier and plot in Figs. 3(g)–3(i) the simulated distributions of the hyperplanes at different steps of training for 3D data classification. The simulation shows that SLAEN's optimized hyperplanes (red circles) have a distance of $d_S = 0.135 \pm 0.056$ to the true hyperplane, i.e., the optimum solutions (hexagrams), as compared to a distance of $d_C = 0.167 \pm 0.073$ for the optimized hyperplanes attained by the classical classifier (red circles) (see Appendix C for simulation results and comparison with experiment).

To investigate the performance of SLAEN and the classical classifier with respect to the signal-to-noise ratio of the data, the error probabilities, under the optimum settings for the VQC and the classical SVMs in the 2D data-classification problem, are measured as the margin ϵ varies in $\{0.2, 0.4, 0.6, 0.8, 1\}$. The results plotted in Fig. 4 show
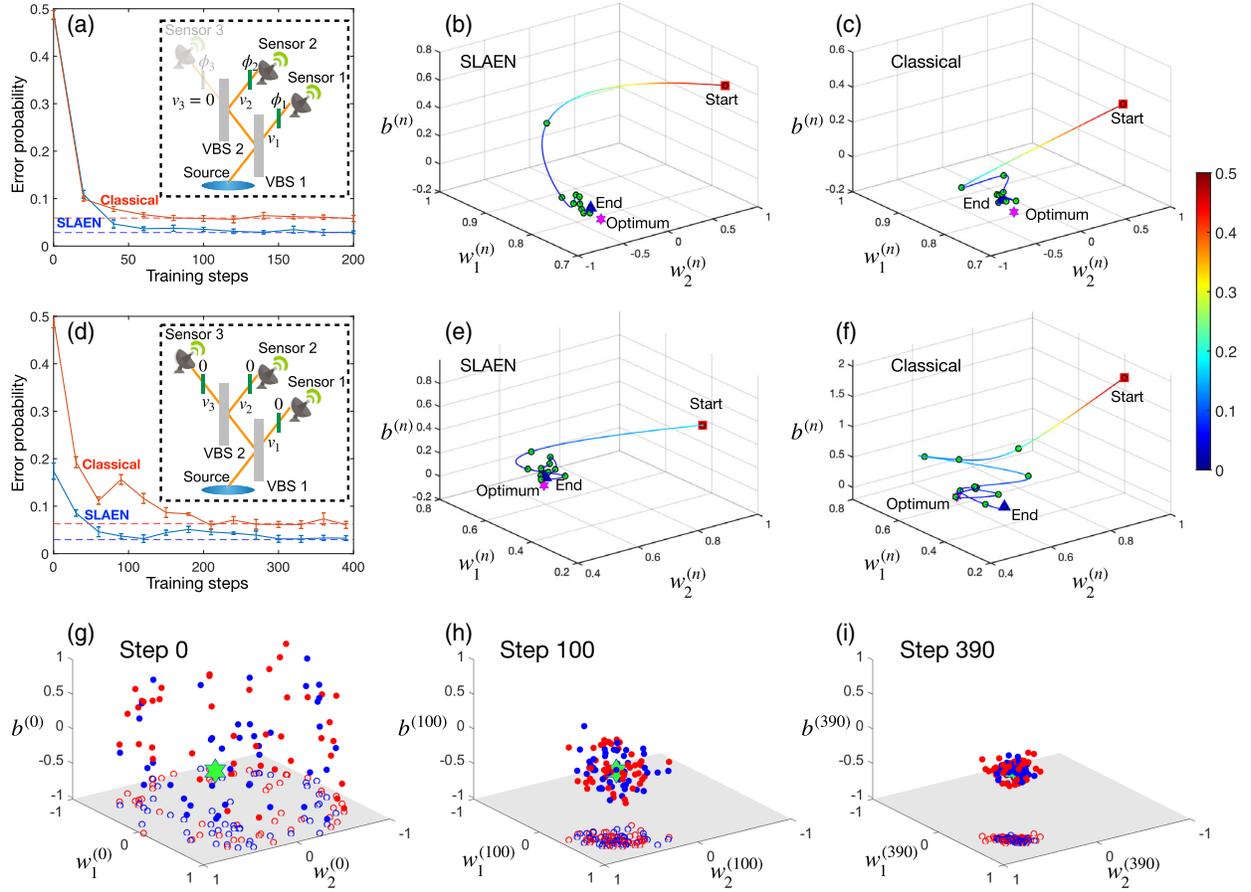
FIG. 3. Experimental results for training SLAEN and classical classifier. Convergence of error probabilities during training for 2D data classification (a) and 3D data classification (d). Blue curves: SLAEN. Red curves: classical classifier. Horizontal dashed lines: expected error probabilities based on true hyperplanes and measurement-noise levels. Error bars: 1 standard deviation of uncertainty derived from five measurements, each with 1000 data points. Insets: VQC parameters being optimized. VBS: variable beam splitter. History of hyperplane $\{(w_1^{(n)}, w_2^{(n)}), b^{(n)}\}$ during training for 2D data classification (b,c) and 3D data classification (e,f). (b,e) SLAEN; (c,f) classical classifier. Red squares: initial hyperplane parameters prior to training. Blue triangles: hyperplane parameters after training. Hexagrams: optimum hyperplane parameters. Color gradients: evolution of error probabilities during training. Green circles: samples of hyperplane parameters at every 20 (30) training steps for 2D (3D) data classification. Curves are obtained from a cubic spline data fitting. Simulated distributions of hyperplane parameters prior to training (g), at Step 100 (h), and at Step 390 (i). Blue filled circles: SLAEN hyperplanes. Red filled circles: classical-classifier hyperplanes. Hexagrams: optimum hyperplanes. Open circles: projected hyperplane parameters onto (w_1, w_2) plane (grey). SLAEN's optimized hyperplanes distribute statistically closer to optimum solutions.

that SLAEN enjoys an error-probability scaling advantage over that of the classical classifier, as manifested in the disparity between the slopes for the two error-probability curves. At $\epsilon = 1$, SLAEN's error probability is more than threefold less than that of the classical classifier.

IV. DISCUSSIONS

The SLAEN theory paper [33] reported an error-probability advantage achieved by an entangled sensor network over that of a sensor network based on separable squeezed states with the same total number of photons, which is verified by the current SLAEN experiment (see Appendix D for details). However, SLAEN's performance has been

primarily benchmarked against classical classifiers that do not use any quantum resources (see Ref. [34] for an in-depth discussion about different types of resources used in distributed quantum sensing). Such a choice is motivated by two main considerations. First, the compared classical classifier represents a common configuration for sensing and data processing. Introducing quantum resources yields a performance enhancement over the existing classical schemes. In the SLAEN experiment, the power of the coherent-state portion is orders of magnitude stronger than that of either the squeezed or the entangled light, similar to the case in the squeezed-light-enhanced Laser Interferometer Gravitational Wave Observatory (LIGO) [35]. In both cases, the squeezed-light power is limited due to experimental capabilities, so it
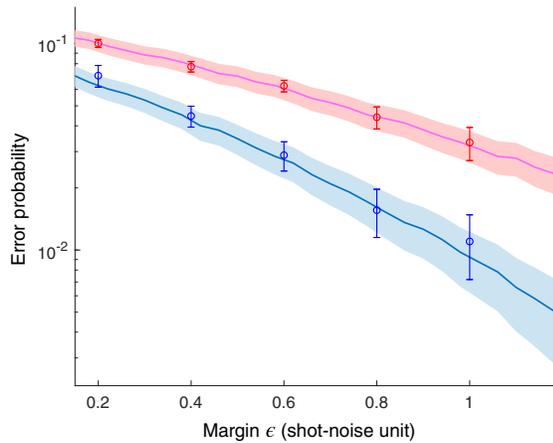
FIG. 4.   Scaling of error probability vs margin of the data set. Blue: SLAEN. Red: classical classifier. Circle: estimated error probability based on five sets of 1000 experimental data points. Solid lines: error probabilities obtained from Monte Carlo simulations. Shades: estimated uncertainty with 1000 samples. SLAEN shows an error-probability scaling advantage over the classical classifier.

barely affects the total optical power employed in sensing. As such, like LIGO, we choose to quantify the quantum advantage as the performance gain over the classical system using the same amount of laser power but not taking advantage of any quantum resources. Second, a complete experimental demonstration of supervised learning based on separable squeezed states requires three independent squeezed-light sources, which places significantly more resource overhead than SLAEN's single squeezed-light source. Hence, SLAEN also enjoys a practical advantage over classical classifiers based on separable squeezed states. It is worth noting that such a practical advantage would be more pronounced when sensors are nearby so that the entanglement distribution loss is low.

Our experiment has implemented an entanglement source trained by supervised learning. The original SLAEN proposal [33], however, also entails reconfigurable measurements. Since homodyne measurements commute with a linear quantum circuit, SLAEN's performance under three homodyne detectors equals that obtained by the variational measurement apparatus considered by Ref. [33]. The current SLAEN protocol only leverages Gaussian sources and measurements, but non-Gaussian resources would potentially improve its performance. Indeed, non-Gaussian measurements have been shown to benefit quantum metrology [36], quantum illumination [37], and entanglement-assisted communication [38]. A variational circuit approach for non-Gaussian entanglement generation and measurements would open a promising route to further enhance the performance.

V. CONCLUSIONS

In conclusion, we have experimentally demonstrated the SLAEN framework for quantum-enhanced data classification. Our work opens a new route for exploiting NISQ hardware to enhance the performance of real-world data-processing tasks. Our current experiment verified SLAEN's quantum advantage in classifying features embedded in rf signals, but SLAEN by itself is a general framework applicable to data-processing problems in other physical domains by appropriately engineering entangled probe states and quantum transducers. The present experiment only demonstrated data classification with linear hyperplanes. To accommodate nonlinear hyperplanes, non-Gaussian entangled probe states [39] and joint quantum measurements [40] would be needed, and the VQC parameter optimizer would also need to be trained to conduct an effective mapping from the SVM hyperplane to the VQC parameters. With these developments, we envisage that SLAEN would create new near-term opportunities in a variety of realms, including distributed big-data processing, navigation, chemical sensing, and biological imaging.

APPENDIX A: OPTIMIZATION ALGORITHM

The simultaneous perturbation stochastic approximation (SPSA) algorithm is used by the classical support-vector machine (SVM) to update the hyperplane in each training step. The SPSA algorithm calculates an approximation of the gradient with only two measurements, w_+ and w_-, of the loss function. This simplicity leads to a significant complexity reduction in the cost optimization. See Algorithm 1 for details.

In the algorithm, d is the dimension of the data set. Here, $d = 2$ for classification problems in a 2D data space, while $d = 3$ for classification problems in a 3D data space. The choices of a, c, A, and γ determine the gain sequences a_k and c_k, which in turn set the learning rates and have a significant impact on the performance of the SPSA algorithm. The parameters used by the classical SVM in our experiment are $a = 1, c = 1, A = 200, \alpha = 0.602$, and $\gamma = 0.1$.

The SPSA algorithm calls a loss function that is in line with the form of the cost function [Eq. (1) of the main text] but allows for an iterative optimization, as defined below:

$$\text{loss}(\boldsymbol{w}, b) = |1 - y^{(n)}(\boldsymbol{w} \cdot \tilde{\boldsymbol{\alpha}}^{(n)} + b)|_+ + \lambda \|\boldsymbol{w}\|^2. \quad \text{(A1)}$$

Algorithm 1. Simultaneous perturbation stochastic approximation (SPSA) [41].

Initialization $a; c; A; \alpha; \gamma; d; N; \mathbf{w}^{(0)}; b^{(0)}$
for $n = 1 : N$ **do**
$\quad a_n = a/(n + A)^\alpha$
$\quad c_n = c/n^\gamma$
$\quad \Delta_w = 2 * \text{round}(\text{rand}(d, 1)) - 1$
$\quad \mathbf{w}_+ = \mathbf{w}^{(n-1)} + c_n * \mathbf{\Delta}_w$
$\quad \mathbf{w}_- = \mathbf{w}^{(n-1)} - c_n * \mathbf{\Delta}_w$
$\quad \Delta_b = 2 * \text{round}(\text{rand}(1, 1)) - 1$
$\quad b_+ = b^{(n-1)} + c_n * \mathbf{\Delta}_b$
$\quad b_- = b^{(n-1)} - c_n * \mathbf{\Delta}_b$
$\quad l_+ = \text{loss}(\mathbf{w}_+, b_+)$
$\quad l_- = \text{loss}(\mathbf{w}_-, b_-)$
$\quad g_w = (l_+ - l_-)/(2 * c_n * \mathbf{\Delta}_w)$
$\quad \mathbf{w}^{(n)} = \mathbf{w}^{(n-1)} - a_n * g_w$
$\quad g_b = (l_+ - l_-)/(2 * c_n * \mathbf{\Delta}_b)$
$\quad b^{(n)} = b^{(n-1)} - a_n * g_b$
end

APPENDIX B: EXPERIMENTAL DETAILS

1. Experimental setup

A detailed experimental setup is shown in Fig. 5. Squeezed light at 1550 nm is generated from an optical parametric amplifier (OPA) cavity where a type-0 periodically poled KTiOPO$_4$ (PPKTP) crystal is pumped by light at 775 nm produced from a second harmonic generation (SHG) cavity. The cavities are locked by the Pound-Drever-Hall technique using 24-MHz sidebands created by phase modulating the 1550-nm pump light prior to the SHG. A small portion of light at 1550 nm modulated at 20 MHz is injected into the OPA cavity and phase locked to the pump light to operate in a parametric amplification regime. In doing so, the squeezed light emitted from the OPA cavity is composed of an effective single-mode squeezed vacuum state residing in the 11-MHz sidebands and a displaced

phase squeezed state at the central spectral mode. Because of the large quadrature displacement at the central spectral mode, it can be well approximated by a classical coherent state. More details about the characterization of our squeezed-light source are enclosed in the Supplemental Material of Ref. [31].

The squeezed light is directed to a VQC composed of two VBSs and three phase shifters, parametrized by $\mathbf{v} \equiv \{v_1, v_2, v_3, \phi_1, \phi_2, \phi_3\}$. Here, v_m is the portion of the power diverted to the mth sensor, satisfying $\sum_{m=1}^3 v_m = 1$, and ϕ_m is the phase shift on the quantum state at the mth sensor. Each VBS comprises a first half-wave plate, a quarter-wave plate, a PM, a second half-wave plate, and a polarizing beam splitter. The power splitting ratio is controlled by applying a dc voltage generated from a computer-controlled data acquisition board (NI PCI 6115). The dc voltage is further amplified by a high-voltage amplifier (Thorlabs HVA 200) with a gain of 20 prior to being applied on the PM. The power portions are determined by

$$v_1 = \frac{1}{2}\left(\sin\left(\frac{E_{s_1}}{V_\pi}\pi\right) + 1\right),$$
$$v_2 = 1 - v_1 - v_3,$$
$$v_3 = \frac{1}{2}\left(\sin\left(\frac{E_{s_2}}{V_\pi}\pi\right) + 1\right)(1 - v_1), \quad \text{(B1)}$$

where E_{s_1} and E_{s_2} are dc voltages applied on PM 1 and PM 2.

After the VBSs, the three-mode entangled probe state, represented by the annihilation operators $\{\hat{b}_1, \hat{b}_2, \hat{b}_3\}$, is diverted to three rf-photonic sensors, each equipped with an EOM driven by a rf signal at a 11-MHz carrier frequency. Because of the phase modulation, a small portion of the coherent state at the central spectral mode is transferred to the 11-MHz sidebands, inducing a phase quadrature displacement. The quadrature displacement at each rf-photonic sensor is equal to [31]
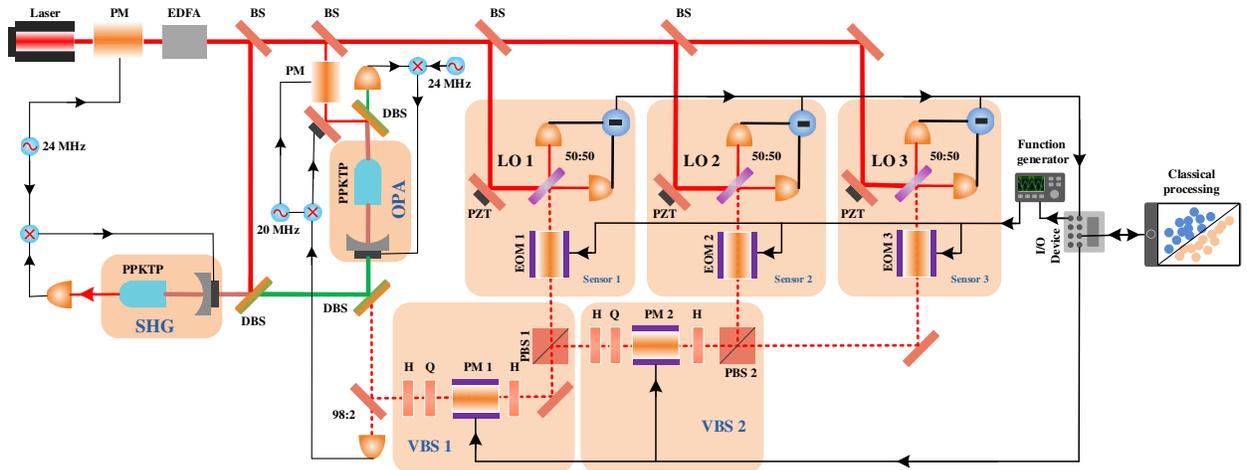
FIG. 5. Detailed experimental diagram.

$$\langle \hat{b}_m \rangle = \alpha_m \simeq i\sqrt{2}\pi g_m a_{c_m} \frac{\gamma E_m}{2V_\pi} \sin(\varphi_m), \qquad \text{(B2)}$$

where $g_m = \pm 1$ is set by a rf signal delay that controls the sign of the displacement. Choosing $g_m = -1$ is equivalent to introducing a π-phase shift on the quantum state at the mth sensor [31], i.e., setting $\phi_m = \pi$ in the VQC parameters. In Eq. (B2), a_{c_m} is the amplitude of the baseband coherent state at the mth sensor. Specifically, $a_{c_m} = \sqrt{v_m}\beta$, where β is the amplitude of the baseband coherent state at the squeezed-light source. Here, V_π is the half-wave voltage of the EOM, and γ describes the conversion from an external electric field E_m to the internal voltage. A more detailed theoretical model for the setup was presented in Ref. [31].

Subsequently, phase-quadrature displacements carried on the quantum light at the three sensors are measured in three balanced homodyne detectors. At each homodyne detector, the quantum light and the local oscillator (LO) first interfere on a 50/50 beam splitter with a characterized interference visibility of 97% and then detected by two photodiodes, each with about 88% quantum efficiency. The difference photocurrent is amplified by a transimpedance amplifier with a gain of 20×10^3 V/A. The dc component of the output voltage signal locks the phase between the LO and the quantum light. The 11-MHz ac component of the voltage signal is demodulated by an electronic mixer, filtered by a 240-kHz low pass filter, and then amplified by a low-noise voltage preamplifier (Stanford Research Systems SR560). The data are acquired by a multifunction I/O device (NI USB-6363) and further processed by a desktop computer in real time. Summing up the measurement data from the three sensors appropriately by $\sum_m \sqrt{v_m} \exp(i\phi_m)\tilde{\alpha}_m$ enables the maximal noise reduction, which is equivalent to the

noise of the squeezed quadrature of the single-mode squeezed state $\hat{b}$ at the source [31]:

$$\text{var}(\hat{b}) = \text{var}\left(\sum_m^M \sqrt{v_m} \exp(i\phi_m)\hat{b}_m \right)$$
$$= \frac{1}{4}\left[\frac{\eta}{(\sqrt{N_s} + \sqrt{N_s + 1})^2} + (1 - \eta) \right], \qquad \text{(B3)}$$

where η is the quantum efficiency at each sensor and N_s is the total photon number of the single-mode squeezed light at the source. In our experiment, $\eta \sim 53\%$ and $N_s \simeq 3.3$. In comparing the summation in Eq. (B3) with that in Eq. (1), it becomes clear that choosing the mapping from the hyperplane parameter w to the VQC setting v to be $w_m = \sqrt{v_m} \exp(i\phi_m)$ minimizes the measurement noise. In the current setup, we measured a 2.9 (3.2) dB noise reduction with the three- (two-) mode entangled state. The characterization of our sensor networks has been reported in Ref. [31].

2. Calibration

a. Calibration of variational quantum circuit

To ensure accurate configuration of the VQC, we first calibrate the power splitting ratio of both VBSs. In calibrating VBS 1, we scan the voltage E_{s_1} applied on PM 1 and measure the transmitted optical power, as plotted in Fig. 6(a). The data are fitted to a sinusoidal function in Eq. (B1), which derives $V_\pi = 606$ V for PM 1. An identical calibration procedure is applied on VBS 2, and we obtain $V_\pi = 606$ V for PM 2.

We then measure the quadrature displacements under different VBS splitting ratios, as a means to test the locking stability between the quantum signal and the LO. To do so,
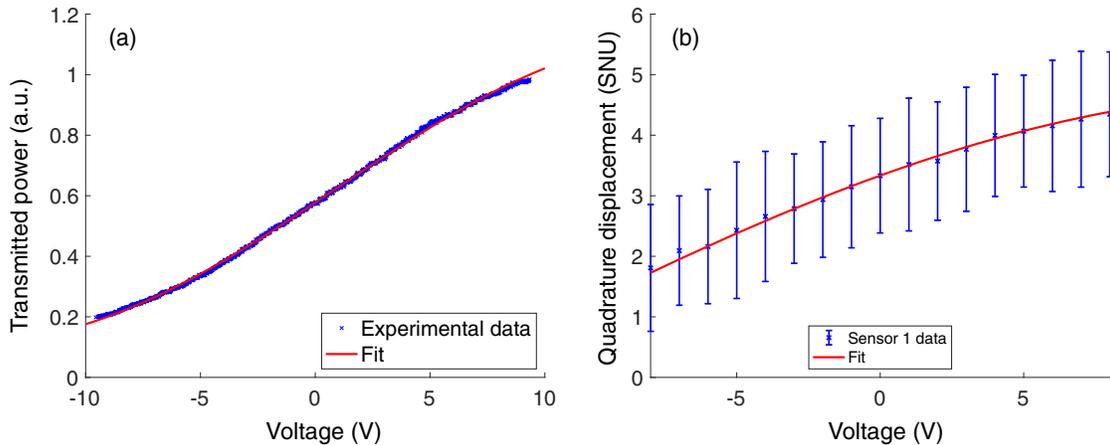
FIG. 6. Calibration of the variable beam splitter. (a) Light power delivered to Sensor 1 from VBS 1 (blue crosses) with respect to the input voltage to the high-voltage amplifier, which is proportional to E_{S_1} in Eq. (B1). Red line: a sinusoidal fit to v_1 in Eq. (B1), up to a scaling factor. (b) Calibration of quadrature displacement introduced by Sensor 1 vs input voltage to the high-voltage amplifier. Blue crosses: homodyne measurement data at Sensor 1. Red solid line: a fit to $\sqrt{v_1}$ in Eq. (B1), up to a scaling factor. Error bars represent the standard deviations of the measurement results, which are determined by the shot-noise level. SNU: shot-noise unit.

while the quantum signal and LO are phase locked, the VBS transmissivity is randomly set to one of 17 values at 30 Hz, subject to the limited bandwidth of the control system. Note that 100 homodyne measurements of quadrature displacement are taken at each transmissivity at a 500-kHz sampling rate. The fitted data are plotted in Fig. 6(b), showing excellent signal stability and agreement with theory in Eq. (B1). The value of the extrapolated V_π is around 612 V, in good agreement with the specification of the EOM. The tunable range for the VBS transmissivity is between 0.07 and 0.93, limited by the maximum output voltage of the high-voltage amplifier (± 200 V). VBS 2 is calibrated in an identical way, deriving a V_π consistent with that of VBS 1. During training, the transmissivity of the VBS is restricted within 0.125 to 0.875 to ensure sufficient light power for phase locking between the quantum signal and the LO.

b. Calibration of rf-photonic transduction

The training data of the rf-field direction (mean-amplitude) classification are prepared by applying phase (amplitude) modulation on the rf signals. Modulations on rf signals are converted to different quadrature displacements by three EOMs. To ensure linearity in the transduction from the amplitude and phase of the rf signals to quadrature displacements, we calibrate the quadrature displacements at each sensor with respect to the modulation voltages that determine the amplitude and phase for the rf signals applied on the EOMs. In the calibration of phase modulation at Sensor 1 shown in Fig. 7(a), as we sweep the modulation voltage on the function generator for the rf signal from -0.5 V to 0.5 V with an increment of 0.1 V, 100 homodyne measurements of the quadrature displacement are recorded

for each modulation voltage at a 500-kHz sampling rate. The distribution of the experimental data on the vertical axis at a given modulation voltage arises from the quantum measurement noise. The fit shows an excellent linear dependence of quadrature displacement vs the modulation voltage. To calibrate the amplitude modulation on the rf signal, we first set the modulation depth to 120% to allow for a sign flip on the rf signal to enable both positive and negative quadrature displacements. We then take 100 homodyne measurements at each modulation voltage at a 500-kHz sampling rate. The experimental data and fit are plotted in Fig. 7(b), showing excellent linear dependence of the measured quadrature displacement with respect to the amplitude of the rf signal. The other two EOMs are calibrated in the same way.

3. Implementation of data classification

a. Training stage for SLAEN

The training stage consists of N steps using randomly produced training data $\{\boldsymbol{E}^{(n)}, \boldsymbol{\varphi}^{(n)}, y^{(n)}\}_{n=1}^{N}$. In the nth training data point, $\boldsymbol{E}^{(n)} \equiv \{E_m^{(n)}\}_{m=1}^{M}$ and $\boldsymbol{\varphi}^{(n)} \equiv \{\varphi_m^{(n)}\}_{m=1}^{M}$ entail, respectively, the probed rf-field amplitudes and phases at the $M = 2$ or $M = 3$ sensors, and $y^{(n)} \in \{-1, +1\}$ is the true label, which can be derived using the true hyperplane $\{\boldsymbol{w}_t, b_t\}$ for the data-classification problem at hand. Each sensor then converts the probed rf field into an internal voltage signal, which in turn drives the EOM to induce a quadrature displacement on the quantum signal,

$$\alpha_m^{(n)} \simeq i\sqrt{2}\pi g_m^{(n)} a_{c_m}^{(n)} \frac{\gamma E_m^{(n)}}{2V_\pi} \sin(\varphi_m^{(n)}), \qquad \text{(B4)}$$

which is similar to Eq. (B2).
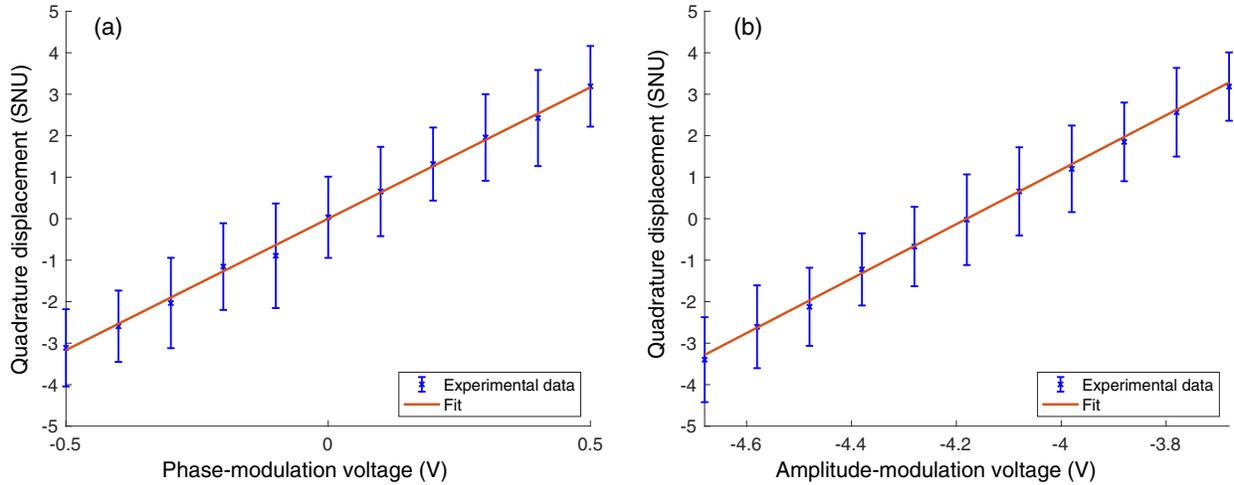
FIG. 7. Calibration of rf-photonic transduction. (a) Calibration of the linearity between the quadrature displacement and the phase of the rf signal proportional to the phase-modulation voltage applied on the function generator. (b) Calibration of the linearity between the quadrature displacement and the amplitude of the rf signal proportional to the amplitude-modulation voltage applied on the function generator. Blue crosses: experimental data acquired from homodyne measurements. Red curve: a linear fit. Error bars represent the standard deviations of the measurement results, which are determined by the shot-noise level. SNU: shot-noise unit.

A technicality associated with the quadrature displacement at the mth sensor is that it depends on both $E_m^{(n)}$ and the amplitude of the baseband coherent light, $a_{c_m}^{(n)} = \sqrt{v_m^{(n)}}\beta$, as shown by Eq. (B4) (see also Ref. [31] for a more detailed description). Our experiment focuses on demonstrating the principle of SLAEN, so, without loss of generality, the displacement's dependence on the baseband light is eliminated by scaling γ by a factor of $1/\sqrt{v_m^{(n)}}$ such that the amount of induced displacement is solely determined by the training data. In our experiment, this is accomplished by applying an extra amplitude modulation that introduces a gain of $1/\sqrt{v_m^{(n)}}$ on the rf signal before it goes to the EOM.

We properly choose γ such that the training data point $\{\mathbf{E}^{(n)}, \boldsymbol{\varphi}^{(n)}, y^{(n)}\}$ leads to random quadrature displacements $\boldsymbol{\alpha}^{(n)}$ at the involved sensors with each displacement value initially following a uniform distribution within $[-4, 4]$ in the shot-noise unit. The signal-to-noise ratio of the training data set is tuned by excluding points within a margin of ϵ to the hyperplane. In doing so, the signal-to-noise ratio is increases as ϵ increases. In the training experiments, $\epsilon = 0.6$ is chosen, comparable to one shot-noise unit.

In the 3D data-classification experiment, amplitude modulations on the rf signals from three function generators prepare the training data. Two dc voltages produced by a multifunction I/O device (NI PCI-6115) are used to configure the two VBSs in the VQC. In the 2D data-classification experiment, phase modulations on the rf signals from two function generators prepare the training data, and one dc voltage generated by the same multifunction I/O device is used to configure VBS 1.

The flowchart of the training process is sketched in Fig. 8. The training starts with an initial hyperplane $\{\mathbf{w}^{(0)}, b^{(0)}\}$ and its corresponding VQC setting $\mathbf{v}^{(0)}$. Here, $b^{(0)}$ is a number stored in the classical SVM algorithm and will be updated during training.

The measurement data at each sensor are collected by a multifunction I/O device (NI USB-6363) operating in an on-demand mode and are then transmitted to a desktop computer on which the classical SVM algorithm runs. In the nth training step, the measurement data $\tilde{\boldsymbol{\alpha}}^{(n)}$ from all sensors, the true label $y^{(n)}$, and the current hyperplane $\{\mathbf{w}^{(n-1)}, b^{(n-1)}\}$ are fed to the SPSA algorithm, which then updates the hyperplane to $\{\mathbf{w}^{(n)}, b^{(n)}\}$, as elaborated in Appendix A. The VQC setting is subsequently updated to $\mathbf{v}^{(n)}$. The next training step starts by adjusting the power splitting ratios of the VBSs by applying two voltages on the PMs based on Eq. (B1) and the calibrated V_π. The new training data are then applied through the EOMs.

During training, a phase shift $\varphi_m^{(n)} = \pi$ needs to be applied to the quantum state $\hat{b}_m$ when $\text{sign}(w_m^{(n)}) = -1$. Experimentally, this is done by flipping the sign of the
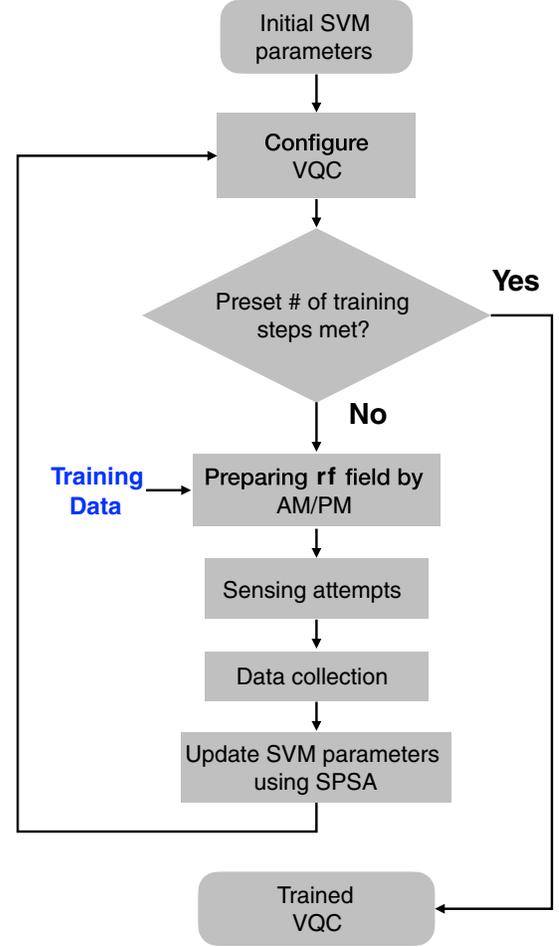
FIG. 8. Flowchart of the training process for SLAEN.

emulated rf-signal amplitude. If the sign of the initial hyperplane, $\text{sign}(w_m^{(0)})$, is different from that of the true hyperplane, $w_m^{(n)}$ will move across zero, which will cause zero optical power to be delivered to the mth sensor such that the phase locking between the quantum signal and the LO breaks down. To avoid this, we restrict the minimum power splitting ratio to $\min(v_m^{(n)}) = 0.125$ so that a sign flip on $w_m^{(n)}$ will be applied whenever $v_m^{(n)}$ hits this boundary. The training iterates 200 steps for the 2D data-classification experiment and 390 steps for the 3D data-classification experiment. The loss function converges to its minimum with the hyperplane $\{\mathbf{w}^\star, b^\star\}$ when training is completed.

b. Utilization stage for SLAEN

In the utilization stage, SLAEN performs data classification on new measurement data $\tilde{\boldsymbol{\alpha}}^{(n)}$, each with an unknown label. The new data follow the same statistical distribution as the training data. To verify the convergence in the training process, we first measure the error probabilities at different training steps with the hyperplane $\{\mathbf{w}^{(k)}, b^{(k)}\}$, where $k \in \{0, 20, 40, \ldots, 160, 180, 200\}$ in
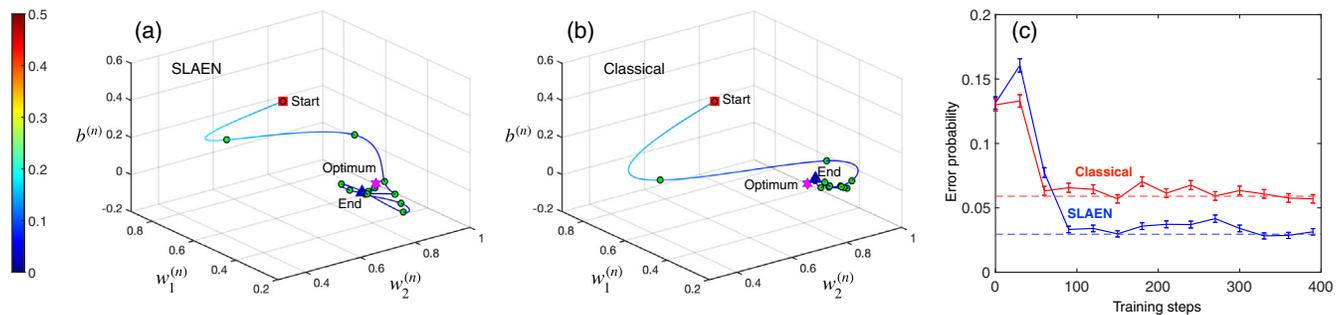
FIG. 9. Experimental optimization of hyperplane parameters for a general data-classification problem. Trajectories of hyperplane parameters for SLAEN (a) and for the classical classifier (b). Red squares: initial hyperplane parameters prior to training. Blue triangles: hyperplane parameters after training. Magenta hexagrams: optimum hyperplane parameters. Color gradients: evolution of the error probabilities during training. Green circles: samples of hyperplane parameters at every 30 training steps. Curves are obtained by a cubic spline data fitting. (c) Error probabilities derived under the hyperplane parameters during training. Blue curve: error probabilities for SLAEN. Red curve: error probabilities for classical classifier.

the 2D data-classification experiment and $k \in \{0, 30, 60, \ldots, 330, 360, 390\}$ in the 3D data-classification experiment. The classical SVM is set to use the hyperplane $\{w^{(k)}, b^{(k)}\}$, and the VQC is configured by the corresponding setting $v^{(k)}$. For each error-probability measurement, 1000 testing data points are applied on the EOMs at a 500-kHz rate by a multifunction I/O device (NI USB-6363), and the measurement data $\tilde{\alpha}^{(n)}$ are synchronously recorded by the same device. A decision is made based on

$$\tilde{y}^{(n)} = \text{sign}(w^{(k)} \cdot \tilde{\alpha}^{(n)} + b^{(k)}). \quad (B5)$$

We then estimate the error probability via $P_E = \sum_{n=1}^{N=1000} |(\tilde{y}^{(n)} - y^{(n)})|/N$.

To verify the scaling of error probability with respect to the signal-to-noise ratio of the data set, the hyperplane parameters and the VQC setting are configured to $\{w^\star, b^\star\}$ and $v^\star$. The error probabilities for the 2D example are measured using data sets with margins $\epsilon \in \{0.2, 0.4, 0.6, 0.8, 1\}$.

c. Training and utilization stages for classical classifier

The measurement noise at different sensors in the classical classifier is independent. As such, the classical classifier can solely be trained in postprocessing carried out by the classical SVM. To perform a direct performance comparison, the training data sets for SLAEN are used to train the classical classifier. The hyperplane $\{w^{(n)}, b^{(n)}\}$ is updated in each training step. The error probabilities at different training steps are measured to validate the convergence. As a comparison, the scaling of the error probabilities for the classical classifier in the 2D example is also measured using the same testing data sets as SLAEN employs.

4. Experiment for general 3D data classification

To show that SLAEN can be trained to tackle general data-classification problems, we randomly choose a true

hyperplane and experimentally train SLAEN and the classical classifier to undertake the classification task. In the experiment, the initial hyperplane is randomly set to $\{w_0 = (0.60, 0.566, 0.566), b_0 = 0.45\}$, and the true hyperplane is $\{w_t = (0.8165, 0.4082, 0.4082), b_t = 0\}$. A training data point is supplied to SLAEN at each of the 390 steps, during which the evolving hyperplane parameters are recorded. As anticipated, the experimental result depicted in Fig. 9(a) shows that the hyperplane parameters move toward the optimum during training, indicating SLAEN's capability of solving general data-classification problems as long as training data are provided. As a comparison, we train the classical classifier over 390 steps with the same training data set used for SLAEN. The evolving hyperplane parameters during training are plotted in Fig. 9(b), showing that the classical classifier can also shift the hyperplane toward the optimum.

With the experimentally measured hyperplane parameters during training, the error probabilities for SLAEN and the classical classifier are derived and plotted in Fig. 9(c). SLAEN possesses a clear error-probability advantage over the classical classifier. Specifically, the error probability of SLAEN is twofold less than that of the classical classifier when both are trained.

APPENDIX C: SIMULATIONS

We have performed Monte Carlo simulations for the training processes of SLAEN and the classical classifier on a classical computer, as a means to verify the qualitative behaviors of the evolving hyperplane parameters and error probabilities during the training experiments. Note that such a training simulation is merely a testing tool and cannot replace the physical training of SLAEN or the classical classifier in their practical applications because the original data $\{E^{(n)}, \varphi^{(n)}\}$ probed by the sensors are, in general, unavailable.

1. Simulation for two-dimensional data classification

The simulation of the training for 2D data classification undergoes 200 steps, each of which consumes a randomly generated data point. The measurement noise for SLAEN and the classifier is also randomly generated, with the correlation between the measurement noise at SLAEN's different sensors accounted for. To facilitate the comparison between the experimental data and the simulation results, Figs. 3(a)–3(c) in the main text are replicated as Figs. 10(a)–10(c) here. Figure 10(d) depicts the simulated convergence of error probabilities for SLAEN (blue) and

the classical classifier (red). In addition, we simulate the evolving hyperplane $\{w_1^{(n)}, w_2^{(n)}, b^{(n)}\}$ during training and plot the results for SLAEN in Fig. 10(e) and for the classical classifier in Fig. 10(f). In comparing the top and middle panels of Fig. 10, excellent qualitative agreement between the experimental data and simulation results is found. Note that since the experimental and simulated measurement results are random, and the SPSA algorithm is stochastic, we only expect a qualitative agreement.

The simulation, in analogy to the experiment, shows that SLAEN's optimized hyperplane (blue triangle) resides
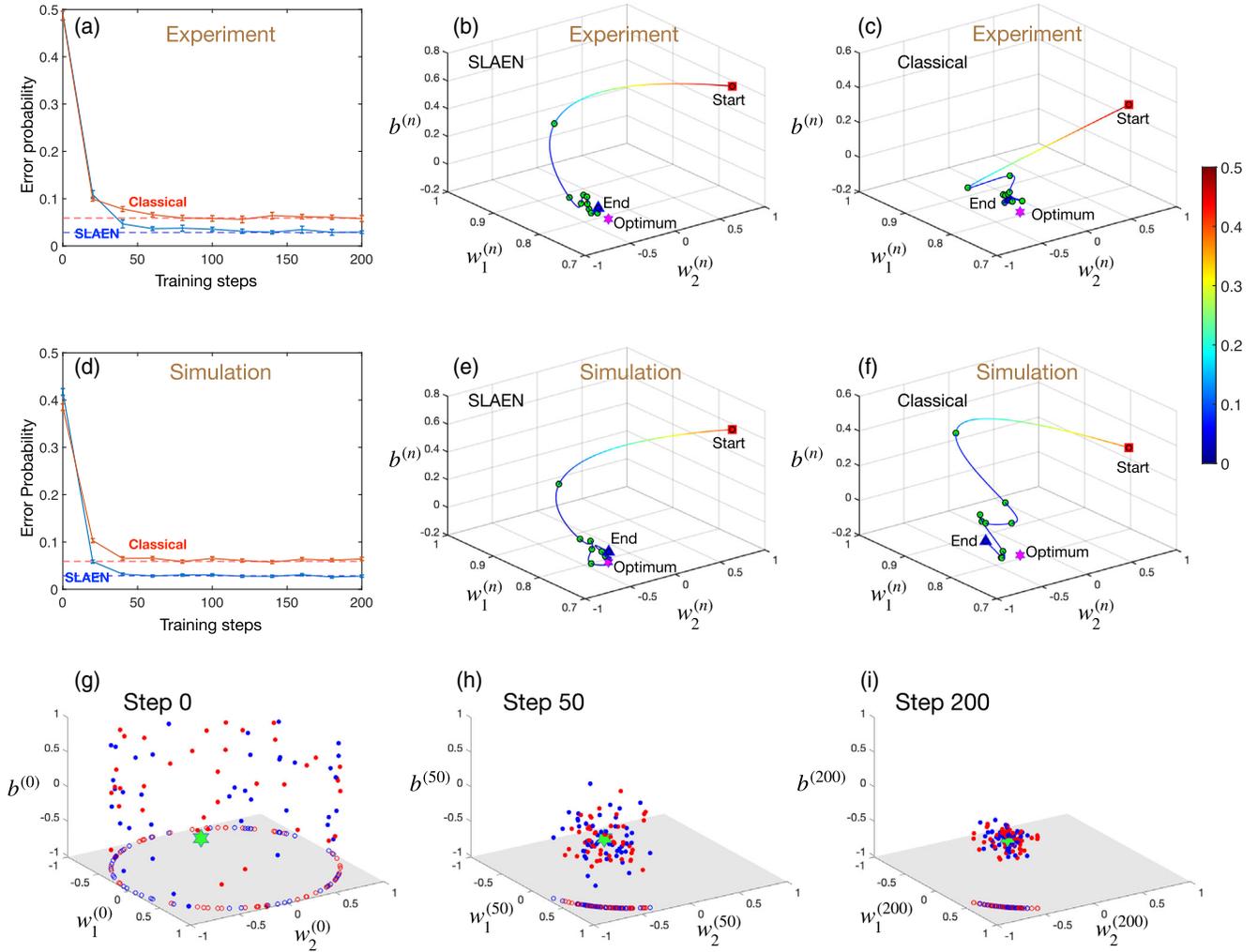
FIG. 10. Comparison between experimental data and simulations in training for 2D data classification. (a,d) Convergence of the error probabilities during training. Red curve: classical classifier. Blue curve: SLAEN. Horizontal dashed lines: expected error probabilities based on true hyperplanes and measurement-noise levels. Error bars represent 1 standard deviation of the uncertainty derived from five measurements or simulations, each with 1000 data points. (b,e) History of hyperplane parameters for SLAEN during training. (d,f) History of hyperplane parameters for classical classifier. Red squares: initial hyperplane parameters prior to training. Blue triangles: hyperplane parameters after training. Magenta hexagrams: true hyperplane parameters, representing the optimum. Color gradients: evolution of the error probabilities during training. Green circles: samples of hyperplane parameters at every 20 training steps. Curves obtained by cubic spline data fitting. Simulated distribution of hyperplane parameters prior to training (g), at Step 50 (h), and at Step 200 (i). Blue filled circles: SLAEN hyperplanes. Red filled circles: classical-classifier hyperplanes. Hexagrams: optimum hyperplanes. Open circles: projected hyperplane parameters onto the (w_1, w_2) face drawn in grey. SLAEN's optimized hyperplanes are distributed statistically closer to the optimum solutions.

closer to the optimum hyperplane (hexagram) than the classical classifier's optimized hyperplane. To investigate whether this is a universal feature for SLAEN, we performed 200 training simulations for both SLAEN and the classical classifier. The initial hyperplane for each training simulation is randomly drawn and is defined as random variables $\{W_S^{(0)}, B_S^{(0)}\}$ for SLAEN and $\{W_C^{(0)}, B_C^{(0)}\}$ for the classical classifier. Figure 10(g) plots the distributions for 50 initial hyperplane parameters in filled circles for both SLAEN (blue) and the classical classifier (red) prior to training. The optimum hyperplane $\{w_t, b_t\}$ is represented by the hexagram. The open circles are the projected hyperplane parameters onto the (w_1, w_2) plane in grey. After 50 training steps, the distributions of the hyperplane $\{W_S^{(50)}, B_S^{(50)}\}$ and $\{W_C^{(50)}, B_C^{(50)}\}$ are drawn in Fig. 10(h), showing that the hyperplane parameters are migrating toward the optimum. The distributions of the hyperplane parameters after 200 training steps are depicted in Fig. 10(i), which shows, qualitatively, that SLAEN's optimized hyperplanes (blue circles) are almost enclosed by the classical classifier's optimized hyperplanes (red circles). This is evidence for SLAEN's enhanced accuracy in seeking the optimum solutions.

To conduct a more quantitative assessment on the convergent behaviors for the hyperplane parameters, we define the distance between SLAEN's hyperplanes and the optimum hyperplane after n training steps as

$$d_S^{(n)} \equiv \left\langle \sqrt{\left(W_S^{(n)} - w_t\right)^2 + \left(B_S^{(n)} - b_t\right)^2} \right\rangle. \quad (C1)$$

The standard deviation of the distance is then defined as

$$\Delta d_S^{(n)} \equiv \sqrt{\left\langle \left[\sqrt{\left(W_S^{(n)} - w_t\right)^2 + \left(B_S^{(n)} - b_t\right)^2} - d_S^{(n)}\right]^2 \right\rangle}. \quad (C2)$$

Likewise, the distance between the classical classifier's hyperplanes and the optimum hyperplane is defined as

$$d_C^{(n)} \equiv \left\langle \sqrt{\left(W_C^{(n)} - w_t\right)^2 + \left(B_C^{(n)} - b_t\right)^2} \right\rangle. \quad (C3)$$

The standard deviation for the classical classifier's distance is then defined as

$$\Delta d_C^{(n)} \equiv \sqrt{\left\langle \left[\sqrt{\left(W_C^{(n)} - w_t\right)^2 + \left(B_C^{(n)} - b_t\right)^2} - d_C^{(n)}\right]^2 \right\rangle}. \quad (C4)$$

The distances at different training steps are plotted in Fig. 11 for SLAEN's hyperplanes (red) and the classical classifier's hyperplanes (blue). The distance for SLAEN's hyperplanes after 200 training steps is $d_S^{(200)} = 0.129 \pm 0.07$, as compared to the classical classifier's $d_C^{(200)} = 0.154 \pm 0.073$. The disparity between the distances at the end of training is highlighted via a zoomed-in view between Step 150 and Step 200 in the inset of Fig. 11.
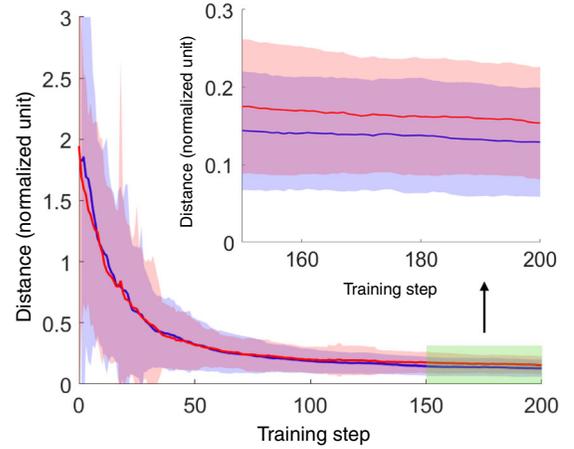
FIG. 11. Distance between the true hyperplane and the optimized hyperplanes at different steps under training for 2D data classification. Solid lines: distance averaged over 200 trajectories for the hyperplane parameters, each with a randomly generated initial hyperplane. Shaded area: standard deviation of the distances of 200 trajectories. Inset: zoom-in view of the distances between training Step 150 and Step 200. Blue: SLAEN. Red: classical classifier.

2. Simulation for three-dimensional data classification

We next simulate the training processes of SLAEN and the classical classifier for 3D data classification. The training for each case takes 390 steps, identical to the number of training steps in the experiment. To facilitate the comparison between experimental data and the simulation results, the plots in Figs. 3(d)–3(f) of the main text are replicated as Figs. 12(a)–12(c) here. The simulated convergence of error probabilities is plotted in Fig. 12(d). Figures 12(e) and 12(f) draw, respectively, the simulated histories of the hyperplane parameters for SLAEN and the classical classifier during training. The qualitative behaviors for the experimental data agree very nicely with those of the simulation results, thereby supporting the validity of the experimental approach.

In addition, we conducted a statistical study of the distances between the hyperplanes and the optimum hyperplane during training for 3D data classification. The distributions of the hyperplane parameters for SLAEN and the classical classifier are plotted in Figs. 3(g)–3(i) of the main text for, respectively, the initial hyperplanes, the hyperplanes after 100 training steps, and the hyperplanes when training is completed. It can be visually observed that SLAEN's optimized hyperplanes are located closer to the optimum hyperplane than the classical classifier's optimized hyperplanes. As a quantitative analysis, the distances vs training step curves for SLAEN and the classical classifier are plotted in Fig. 13. Akin to Fig. 11, SLAEN enables a reduced distance between its optimized hyperplanes and the optimum hyperplane. This is a consequence of the entanglement-enabled measurement-noise reduction
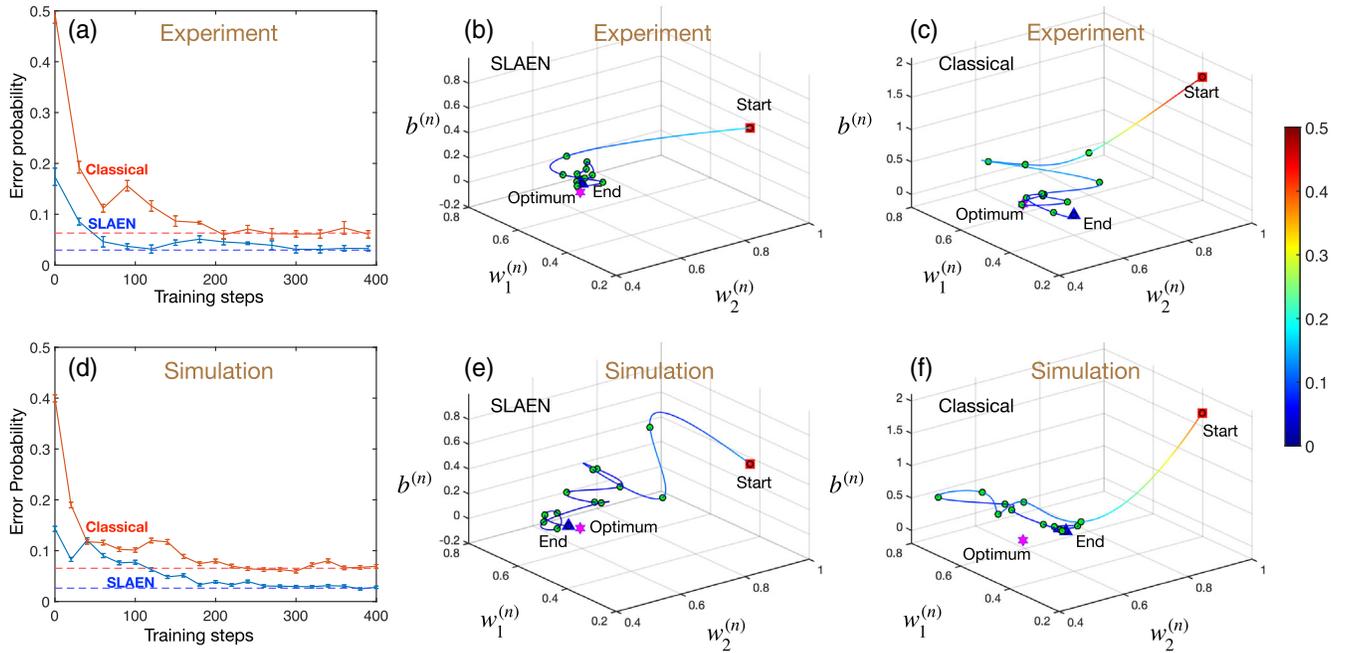
FIG. 12. Comparison between experimental data and simulation results of training for 3D data classification. (a,d) Convergence of the error probabilities during training. Red curve: classical classifier. Blue curve: SLAEN. Horizontal dashed lines: expected error probabilities based on true hyperplanes and measurement-noise levels. Error bars represent 1 standard deviation of the uncertainty derived from five measurements or simulations, each with 1000 data points. (b,e) History of hyperplane parameters for SLAEN during training. (d,f) History of hyperplane parameters for classical classifier during training. Red squares: initial hyperplane parameters prior to training. Blue triangles: hyperplane parameters after training. Magenta hexagrams: optimum hyperplane parameters. Color gradients: evolution of the error probabilities during training. Green circles: samples of hyperplane parameters at every 30 training steps. Curves are obtained by cubic spline data fitting.
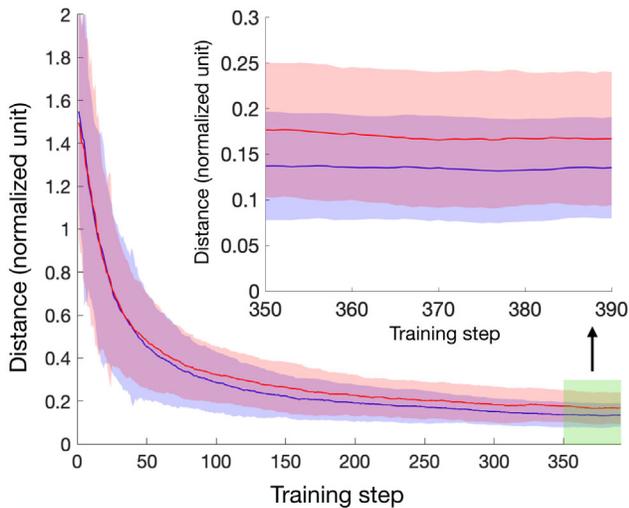
FIG. 13. Distance between the true hyperplane and the optimized hyperplanes at different steps under training for 3D data classification. Solid lines: distance averaged over 200 trajectories for the hyperplane parameters, each with a randomly generated initial hyperplane. Shaded area: standard deviation of the distances of 200 trajectories. Inset: zoomed-in view of the distances between training Step 350 and 390. Blue: SLAEN. Red: classical classifier.

mechanism that SLAEN harnesses. The inset of Fig. 13 is a zoomed-in view of the distances for SLAEN and the classical classifier near the end of training. After 390 training steps, we define SLAEN's optimized distance as $d_S \equiv d_S^{(390)} = 0.135 \pm 0.056$ and the classical classifier's optimized distance as $d_C \equiv d_C^{(390)} = 0.167 \pm 0.073$. Both d_S and d_C are reported in the main text.

APPENDIX D: DATA CLASSIFICATION USING SEPARABLE SQUEEZED STATES

The original SLAEN theory paper [33] showed that the performance of data-processing tasks undertaken by an entangled sensor network is superior to that of a sensor network based on separable squeezed states that have the same total photon number as the entangled state. In this Appendix, we show, in simulation and by experimental data, that our SLAEN experiment achieves an advantage over a sensor network based on separable squeezed states in data-classification tasks, subject to a photon-number constraint. Specifically, our simulation shows that data classification based on separable squeezed states has a larger error probability than that of our SLAEN experiment. We also experimentally show that the quantum noise of a sensor network with separable squeezed states is higher

than that of SLAEN, thereby supporting the SLAEN's claimed advantage over supervised learning based on separable squeezed states. Finally, we present the motivation behind the main text's focus on a performance comparison between SLAEN and classical classifiers based on coherent states.

1. Simulation for 3D data classification using separable squeezed states

We simulate the training process of average rf-field amplitude classification undertaken by a three-node sensor network based on separable squeezed states. The total photon number of the separable squeezed states is set to be the same as that of the entangled states in the SLAEN experiment. The evolution and convergence of the error probabilities during the training process are plotted in Fig. 14. In the simulation, the initial hyperplane is set to $\{\boldsymbol{w}_0 = (0.9044, 0.3152, 0.2876), b_0 = 0.53\}$, identical to that of the training in the SLAEN experiment. As a comparison, we experimentally measure the error-probability evolution in SLAEN by taking ten more measurements based on the same experimental setting as what is used to produce Fig. 12(a). Our SLAEN experiment shows an error-probability advantage of about 13% over that of a simulated sensor network based on separable squeezed states.

2. Experimental noise calibrations

A complete demonstration of a three-node sensor network with separable squeezed states requires three independent squeezed-light sources, which places a significant resource overhead. Instead, we calibrate the quantum noise of a sensor network with separable squeezed states using a time-domain multiplexing approach introduced by Ref. [30]. We first set the mean photon number of our squeezed-light source to that of a separable squeezed state at a single sensor. We then take three samples in the time domain to emulate the independent quantum noise at three sensor nodes. The histogram of the averaged homodyne data is plotted in Fig. 15 and fitted with a normalized Gaussian probability density function. Since the measured noise variance of the separable sensor network is about 11.7% higher than that of SLAEN, it is anticipated that the error probability of SLAEN beats that of a sensor network based on separable squeezed states.

3. Performance comparison

Quantum metrology studies how nonclassical resources such as squeezed light and entanglement can be utilized in a measurement system to enable a performance advantage over systems based on classical resources. Such a performance gain in sensing underpins SLAEN's error-probability advantage over separable sensor networks. In many practical optical sensing systems such as the LIGO, the usable power of the classical laser light is limited due to, e.g., thermal effects, photon radiation-pressure-induced torques, and parametric instabilities that cause adverse effects on the system performance [35]. Nonclassical squeezed light is then injected into the system to further improve the measurement sensitivity. In such a scenario, the measurement sensitivity achieved by a classical laser at a given power level is defined as the standard quantum limit (SQL).

FIG. 14. Convergence of error probability. Blue curve: simulation of training process based on separable squeezed states. Error bars represent 1 standard deviation of the uncertainty derived from 60 000 simulated data points. Red curve: SLAEN experiment. Error bars represent 1 standard deviation of the uncertainty derived from 10 measurements. Horizontal dashed lines: expected error probabilities based on true hyperplanes and measurement-noise levels.

FIG. 15. Histograms of homodyne data. Red bins: separable squeezed states. Blue bins: entangled state. Histograms are normalized to probability mass functions and fitted with Gaussian probability density functions. Red curve: theory fit for separable squeezed states. Blue curve: theory fit for entangled state. Black curve: standard quantum limit.

Surpassing the SQL using nonclassical resources demonstrates a quantum advantage enabled by quantum metrology. In our experiment, SLAEN's performance is compared with that of a classical classifier based on laser light, i.e., coherent states. The error probabilities for a classical classifier are measured at a given laser power level. SLAEN's error probabilities are then measured at the same laser power level while entanglement is distributed and shared by the sensors. In the SLAEN experiment, the calibration gives a total photon number of the entangled state of $N_S = 3.3$ and the quantum efficiency of $\eta = 0.53$ at each sensor. In a conceived three-node sensor network based on separable squeezed states, the mean photon number of a separable squeezed state at each sensor will be 1.1, so the total photon number matches that of the entangled state. We can then estimate a noise reduction of 2.57 dB below SQL at the same quantum efficiency at each sensor as the SLAEN experiment. The squeezed state residing at the 11-MHz sidebands is at a power level of tens of picowatts, while in the experiment, most photons at each sensor originate from the strong ($\sim 50\ \mu$W) coherent state at a central wavelength of 1550 nm. Given the close to 6-orders-of-magnitude power disparity between the strong coherent state and the quantum states at the sidebands, a separable sensor network based on either coherent states or separable squeezed states employs nearly identical optical power as an entangled sensor network. This situation is in analogy to LIGO, in which the overall optical power remains nearly unchanged despite the injection of squeezed light into the interferometer.

---

[1] G. Carleo and M. Troyer, *Solving the Quantum Many-Body Problem with Artificial Neural Networks*, Science **355,** 602 (2017).

[2] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, *Quantum Machine Learning*, Nature (London) **549,** 195 (2017).

[3] X. Gao, Z.-Y. Zhang, and L.-M. Duan, *A Quantum Machine Learning Algorithm Based on Generative Models*, Sci. Adv. **4,** eaat9004 (2018).

[4] 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 (London) **567,** 209 (2019).

[5] M. Schuld and N. Killoran, *Quantum Machine Learning in Feature Hilbert Spaces*, Phys. Rev. Lett. **122,** 040504 (2019).

[6] 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).

[7] P. Rebentrost, M. Mohseni, and S. Lloyd, *Quantum Support Vector Machine for Big Data Classification*, Phys. Rev. Lett. **113,** 130503 (2014).

[8] R. Schützhold, *Pattern Recognition on a Quantum Computer*, Phys. Rev. A **67,** 062311 (2003).

[9] N. Moll, P. Barkoutsos, L. S. Bishop, J. M. Chow, A. Cross, D. J. Egger, S. Filipp, A. Fuhrer, J. M. Gambetta, M. Ganzhorn *et al.*, *Quantum Optimization Using Variational Algorithms on Near-Term Quantum Devices*, Quantum Sci. Technol. **3,** 030503 (2018).

[10] J. Preskill, *Quantum Computing in the NISQ Era and Beyond*, Quantum **2,** 79 (2018).

[11] J. Wang, S. Paesani, Y. Ding, R. Santagati, P. Skrzypczyk, A. Salavrakos, J. Tura, R. Augusiak, L. Mančinska, D. Bacco *et al.*, *Multidimensional Quantum Entanglement with Large-Scale Integrated Optics*, Science **360,** 285 (2018).

[12] D. Zhu, N. M. Linke, M. Benedetti, K. A. Landsman, N. H. Nguyen, C. H. Alderete, A. Perdomo-Ortiz, N. Korda, A. Garfoot, C. Brecque *et al.*, *Training of Quantum Circuits on a Hybrid Quantum Computer*, Sci. Adv. **5,** eaaw9918 (2019).

[13] E. Farhi, J. Goldstone, and S. Gutmann, *A Quantum Approximate Optimization Algorithm*, arXiv:1411.4028.

[14] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta, *Hardware-Efficient Variational Quantum Eigensolver for Small Molecules and Quantum Magnets*, Nature (London) **549,** 242 (2017).

[15] J. J. Meyer, J. Borregaard, and J. Eisert, *A Variational Toolbox for Quantum Multi-parameter Estimation*, arXiv:2006.06303.

[16] Y. Nam, J.-S. Chen, N. C. Pisenti, K. Wright, C. Delaney, D. Maslov, K. R. Brown, S. Allen, J. M. Amini, J. Apisdorf *et al.*, *Ground-State Energy Estimation of the Water Molecule on a Trapped-Ion Quantum Computer*, npj Quantum Inf. **6,** 33 (2020).

[17] J. Carolan, M. Mohseni, J. P. Olson, M. Prabhu, C. Chen, D. Bunandar, M. Y. Niu, N. C. Harris, F. N. C. Wong, M. Hochberg *et al.*, *Variational Quantum Unsampling on a Quantum Photonic Processor*, Nat. Phys. **16,** 322 (2020).

[18] J. Otterbach, R. Manenti, N. Alidoust, A. Bestwick, M. Block, B. Bloom, S. Caldwell, N. Didier, E. S. Fried, S. Hong *et al.*, *Unsupervised Machine Learning on a Hybrid Quantum Computer*, arXiv:1712.05771.

[19] E. Farhi, J. Goldstone, S. Gutmann, and H. Neven, *Quantum Algorithms for Fixed Qubit Architectures*, arXiv:1703.06199.

[20] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, *ImageNet: A Large-Scale Hierarchical Image Database*, in *2009 IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2009), pp. 248–255, https://doi.org/10.1109/CVPR.2009.5206848.

[21] A. J. Fenn, D. H. Temme, W. P. Delaney, and W. E. Courtney, *The Development of Phased-Array Radar Technology*, Lincoln Lab. J. **12,** 321 (2000).

[22] M. Baaske and F. Vollmer, *Optical Resonator Biosensors: Molecular Diagnostic and Nanoparticle Detection on an Integrated Platform*, Chem. Phys. Chem. **13,** 427 (2012).

[23] Q. Zhuang, Z. Zhang, and J. H. Shapiro, *Distributed Quantum Sensing Using Continuous-Variable Multipartite Entanglement*, Phys. Rev. A **97,** 032329 (2018).

[24] W. Ge, K. Jacobs, Z. Eldredge, A. V. Gorshkov, and M. Foss-Feig, *Distributed Quantum Metrology with Linear Networks and Separable Inputs*, Phys. Rev. Lett. **121,** 043604 (2018).

[25] T. J. Proctor, P. A. Knott, and J. A. Dunningham, *Multiparameter Estimation in Networked Quantum Sensors*, Phys. Rev. Lett. **120**, 080501 (2018).

[26] K. Qian, Z. Eldredge, W. Ge, G. Pagano, C. Monroe, J. V. Porto, and A. V. Gorshkov, *Heisenberg-Scaling Measurement Protocol for Analytic Functions with Quantum Sensor Networks*, Phys. Rev. A **100**, 042304 (2019).

[27] D. Gatto, P. Facchi, F. A. Narducci, and V. Tamma, *Distributed Quantum Metrology with a Single Squeezed-Vacuum Source*, Phys. Rev. Research **1**, 032024(R) (2019).

[28] C. Oh, C. Lee, S. H. Lie, and H. Jeong, *Optimal Distributed Quantum Sensing Using Gaussian States*, Phys. Rev. Research **2**, 023030 (2020).

[29] M. R. Grace, C. N. Gagatsos, and S. Guha, *Entanglement Enhanced Estimation of Parameter Embedded in Multiple Phases*, arXiv:2004.04152.

[30] X. Guo, C. R. Breum, J. Borregaard, S. Izumi, M. V. Larsen, T. Gehring, M. Christandl, J. S. Neergaard-Nielsen, and U. L. Andersen, *Distributed Quantum Sensing in a Continuous-Variable Entangled Network*, Nat. Phys. **16**, 281 (2020).

[31] Y. Xia, W. Li, W. Clark, D. Hart, Q. Zhuang, and Z. Zhang, *Demonstration of a Reconfigurable Entangled Radio-Frequency Photonic Sensor Network*, Phys. Rev. Lett. **124**, 150502 (2020).

[32] L.-Z. Liu, Y.-Z. Zhang, Z.-D. Li, R. Zhang, X.-F. Yin, Y.-Y. Fei, L. Li, N.-L. Liu, F. Xu, Y.-A. Chen *et al.*, *Distributed Quantum Phase Estimation with Entangled Photons*, Nat. Photonics **15**, 137, (2021).

[33] Q. Zhuang and Z. Zhang, *Physical-Layer Supervised Learning Assisted by an Entangled Sensor Network*, Phys. Rev. X **9**, 041023 (2019).

[34] Z. Zhang and Q. Zhuang, *Distributed Quantum Sensing*, Quantum Sci. Techn. (to be published).

[35] M. Tse, H. Yu, N. Kijbunchoo, A. Fernandez-Galiana, P. Dupej, L. Barsotti, C. Blair, D. Brown, S. Dwyer, A. Effler *et al.*, *Quantum-Enhanced Advanced LIGO Detectors in the Era of Gravitational-Wave Astronomy*, Phys. Rev. Lett. **123**, 231107 (2019).

[36] K. Jiang, C. J. Brignac, Y. Weng, M. B. Kim, H. Lee, and J. P. Dowling, *Strategies for Choosing Path-Entangled Number States for Optimal Robust Quantum-Optical Metrology in the Presence of Loss*, Phys. Rev. A **86**, 013826 (2012).

[37] Z. Zhang, S. Mouradian, F. N. Wong, and J. H. Shapiro, *Entanglement-Enhanced Sensing in a Lossy and Noisy Environment*, Phys. Rev. Lett. **114**, 110506 (2015).

[38] S. Hao, H. Shi, W. Li, J. H. Shapiro, Q. Zhuang, and Z. Zhang, *Entanglement-Assisted Communication Surpassing the Ultimate Classical Capacity*, arXiv:2101.07482.

[39] A. Ourjoumtsev, H. Jeong, R. Tualle-Brouri, and P. Grangier, *Generation of Optical "Schrödinger Cats" from Photon Number States*, Nature (London) **448**, 784 (2007).

[40] Q. Zhuang, Z. Zhang, and J. H. Shapiro, *Optimum Mixed-State Discrimination for Noisy Entanglement-Enhanced Sensing*, Phys. Rev. Lett. **118**, 040801 (2017).

[41] J. C. Spall, *Implementation of the Simultaneous Perturbation Algorithm for Stochastic Optimization*, IEEE Trans. Aerosp. Electron. Syst. **19**, 482 (1998).