ARTICLE OPEN

Transforming Bell's inequalities into state classifiers with machine learning

Yue-Chi Ma[1] and Man-Hong Yung[1,2,3]

In quantum information science, a major challenge is to look for an efficient means for classifying quantum states. An attractive proposal is to utilize Bell's inequality as an entanglement witness, for classifying entangled state. The problem is that entanglement is necessary but not sufficient for violating Bell's inequalities, making these inequalities unreliable in state classification. Furthermore, in general, classifying the separability of states, even for only few qubits, is resource-consuming. Here we look for alternative solutions with the methods of machine learning, by constructing neural networks that are capable of simultaneously encoding convex sets of multiple entanglement witness inequalities. The simulation results indicated that these transformed Bell-type classifiers can perform significantly better than the original Bell's inequalities in classifying entangled states. We further extended our analysis to classify quantum states into multiple species through machine learning. These results not only provide an interpretation of neural network as quantum state classifier, but also confirm that neural networks can be a valuable tool for quantum information processing.

INTRODUCTION

Quantum machine learning is an emerging field of research in the intersection between quantum physics and machine learning, which has profoundly changed the way we interact with data. It represents a new paradigm of processing information, which, at the fundamental level, is still governed by the laws of quantum mechanics. In addition, there is also a real "demand" of using advanced data-processing techniques for gate-fidelity benchmarking and data analysis for the state-of-the art quantum experiments. Therefore, understanding the connection between quantum information science and machine learning is a matter of fundamental and practical interest.

So far, there are several ways where research in quantum machine learning has become fruitful. One way is to design quantum algorithms to speed up classical machine learning.[1,2] On the other hand, the other approach in quantum machine learning is to apply machine-learning methods to study problems in quantum physics and quantum information science. In particular, classical machine-learning methods[3] have been applied to many-body,[4–11] superconducting,[12] bosonic[13], and electronic[14] systems. Furthermore, machine-learning can also be applied to the problem of state preparation,[15] tomography,[10,16] experiments searching.[17] Beyond quantum information science, machine learning also finds applications in particle physics,[18] electronic structure of molecules,[19,20] and gravitational physics.[21] Furthermore, there are many classical methods in machine learning inspired by ideas in physics.[22]

In this work, we are interested in applications of supervised machine learning to the problem of quantum-state classification,[23] which is a generalization of pattern recognition in learning theory. Supervised (machine) learning refers to a set of methods where both data and the corresponding output (called label) are provided as the input. In the classical setting of pattern recognition, we are given a training set $\mathcal{S}$ containing paired values,

$$\mathcal{S} = \{(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), (\boldsymbol{x}_3, y_3), \dots \}, \tag{1}$$

where $\boldsymbol{x}_i$ is a data point and $y_i \in \{0, 1\}$ is a pre-determined label for $\boldsymbol{x}_i$. Based on the training set, the problem of pattern recognition is to construct a low-error classifier (or predictor), in the form of a function, $f : \boldsymbol{x} \to y$, for predicting the labels of new data. The quantum extension of this problem is to replace the data points $\boldsymbol{x}_i$ with density matrices of quantum states ρ_i. Specifically, a quantum state classifier outputs a "label" associated with the state, for example, "entangled" or " unentangled".

Technically, we employ artificial neural networks (ANN)[24] as our machine-learning method. The architecture of ANN shares a similarity with the structure of biological neural networks, which contains a collection of basic units called "artificial neurons". As shown in Fig. 2b, the simplest neural network consists of linear connections and non-linear output. The network can be improved by inserting a hidden layer, as depicted in Fig. 2c.

CHSH inequality

To get started, let us consider an ensemble of quantum states ρ of n qubits; the method is also applicable for qudit systems. Recall that a quantum state is (fully) separable if and only if it can be expressed as a convex combination of product states, i.e.,

$$\rho_{\text{sep}} = \sum_i p_i \rho_i^1 \otimes \rho_i^2 \otimes \dots \otimes \rho_i^n, \tag{2}$$

[1]Center for Quantum Information, Institute for Interdisciplinary Information Sciences, Tsinghua University, Beijing 100084, China; [2]Shenzhen Institute for Quantum Science and Engineering and Department of Physics, Southern University of Science and Technology, Shenzhen 518055, China and [3]Shenzhen Key Laboratory of Quantum Science and Engineering, Southern University of Science and Technology, Shenzhen 518055, China
Correspondence: Y.-C. Ma (mayc13@mails.tsinghua.edu.cn) or M-H. Yung (yung@sustc.edu.cn)

for $0 \leq p_i \leq 1$ and $\sum_i p_i = 1$. Otherwise, the quantum state is entangled.

Entanglement is necessary for a violation of the Bell's inequalities,[25] e.g., the CHSH (Clauser-Horne- Shimony-Holt) inequality,[26]

$$|\langle ab \rangle - \langle ab' \rangle + \langle a'b \rangle + \langle a'b' \rangle|2, \tag{3}$$

where $\langle \cdot \rangle$ represents expectation, $\{a, a'\}$ and $\{b, b'\}$ are the detector settings of parties A and B respectively that take only two values ± 1 (see Fig. 2a). Furthermore,

$$\hat{n} = n_1 \sigma_x + n_2 \sigma_y + n_3 \sigma_z, \tag{4}$$

$\hat{n} \in \{a, a', b, b'\}$, and $\sigma_{x,y,z}$ are the Pauli matrices.

Quantum states violating the CHSH inequality can be labelled as "entangled". However, CHSH inequalities cannot be employed as a reliable tool for entanglement detection. There are two reasons. First, there exist entangled states not violating the Bell's inequalities. To be more specific, the maximally-entangled state, such as

$$|\psi_-\rangle = (|00\rangle - |11\rangle)/\sqrt{2}, \tag{5}$$

for a pair of qubits, can maximally violate the CHSH inequality.[25] However, this tool fails under the circumstances of noise, in the form of a quantum channel. After passing through a depolarizing channel,[27] the resulting state,

$$\rho = p|\psi_-\rangle\langle\psi_-| + (1-p)I/4, \tag{6}$$

where $0 \leq p \leq 1$, violates the CHSH inequality only if $p > 1/\sqrt{2} \simeq 0.707$.[28] However, the state is entangled when $p > 1/3 \simeq 0.333$.[28]

Another reason is that the measurement angles depends on the quantum state. For example, if we choose fixed measurement angles with the following CHSH operator,

$$\prod_{\text{CHSH}} \equiv a_0 b_0 - a_0 b'_0 + a'_0 b_0 + a'_0 b'_0, \tag{7}$$

where $a_0 = \sigma_z$, $a'_0 = \sigma_x$, $b_0 = (\sigma_x - \sigma_z)/\sqrt{2}$, $b'_0 = (\sigma_x + \sigma_z)/\sqrt{2}$, then for any given quantum state of the form,

$$|\psi_{\theta,\phi}\rangle = \cos(\theta/2)|00\rangle + e^{i\phi}\sin(\theta/2)|11\rangle, \tag{8}$$

we have the expectation value,

$$\langle\psi_{\theta,\phi}| \prod_{\text{CHSH}} |\psi_{\theta,\phi}\rangle = \sqrt{2}(\sin\theta\cos\phi - 1), \tag{9}$$

which is equal to $-2\sqrt{2}$ when $\theta = \pi/2$, and $\phi = \pi$, i.e., when $|\psi_{\theta,\phi}\rangle = |\psi_-\rangle$. For a different value of ϕ, e.g., $\phi = \pi/2$, the resulting quantum state can no longer be used to violate this particular CHSH inequality. Therefore, in general, single original CHSH inequality cannot be employed as a reliable tool for detecting quantum entanglement for given quantum states.

RESULTS

In this work, we focus on the task of classifying entangled or separable states, but the method can also be extended for other physical properties. The main challenges include the following:

1. Obtaining full information for a given quantum state becomes resource consuming as the number of qubits increases.
2. It is known to be computational hard (more precisely, NP-hard)[29] in general even if all the information about the states is given.

For the 1st point, instead of full information (e.g., from quantum tomography), we aim at constructing a set of quantum-state classifiers which can reliably output a correct label of a given quantum state in an ensemble, *using only partial information* (i.e., a few observables) about the state. Our strategy is motivated by the development of Bell's inequalities,[25] which was originally used to exclude incompatible classical theories from a few measurement results performed non-locally.

Our strategy is to "transform" Bell's inequalities into a reliable entanglement-separable states classifier. However, the non-locality aspect of Bell's inequalities is not relevant to the construction of our classifier, although we can follow the same experimental setting for an implementation of our proposal.

Here the transformation process involves two levels. First, we ask the following question: "given the same measurement setting, is it possible to optimize the coefficients of CHSH inequality for a better performance, compared with the values $(1, -1, 1, 1, 2)$ employed in the standard CHSH inequality for specific states?" We shall see that the answer to this question is positive. This optimization is linear, in the sense that it gives an optimization function containing a linear combination of the observables as input.

At the second level, instead of linear optimization, we include a hidden layer in the neural network, making it a non-linear optimization process, and at the same time, allow the measurement angles to be varied randomly. As we shall see, in this way, the performance of classifier can be enhanced significantly, relative to the first level. The process can be considered as encoding multiple variants of CHSH inequalities in a neural network. In other words, we are effectively applying many entanglement witnesses simultaneously.

For the 2nd point, we ask a question: "is it possible to construct a universal state classifier for detecting quantum entanglement?" If possible, this would be a valuable tools for many tasks in quantum information theory. However, the challenge is to find a reliable way for labelling the quantum states in the training set. For a pair of qubits, this is possible by using the PPT (positive partial transpose) criterion. We have constructed such a universal state classifier for a pair of qubits; we find that the performance depends heavily on the testing sets; the major source of error comes from the data near the boundary between entangled and separable states.

Later, we shall argue that our ANN architecture is generic for entanglement detection. It depends on the fact that any entangled state is detectable by at least one witness, and multiple "witness inequalities" can be encoded in our model. Our machine-learning method is then applied to several different scenarios of entangled-separable states classification. As an extension, we consider the systems of three qubits. There, the structure is more complicated than two qubits. Compared with PPT, our model are capable to detect entangled states on which PPT cannot detect. Compared with quantum state tomography, the required resources for classifying these quantum states can be reduced in our model.

In the Supplementary Materials, we also construct a quantum state classifier that can identify four types of states, including three types of entangled and one type of fully-separable states, with again only partial information. Furthermore, we have also considered an ensembles of four-qubit systems. We train and analyze the performance of state classifiers in terms of three groups of quantum states, including entangled, separable ones and the states without a correct label. We also provide an example to show our model can be applied for many-qubit systems with significantly reduced computational resources.

Optimizing CHSH operator with machine learning

In this work, we consider two types of machine learning predictors (Fig. 1a) to classify different types of quantum ensembles, namely

(i) tomographic predictors, and
(ii) Bell-like predictors.

Tomographic predictors make use of all information of a given quantum state and is used to benchmark the performance of Bell-like predictors, which employs a subset of non-orthogonal measurements setting. For example, for a pair of qubits, the
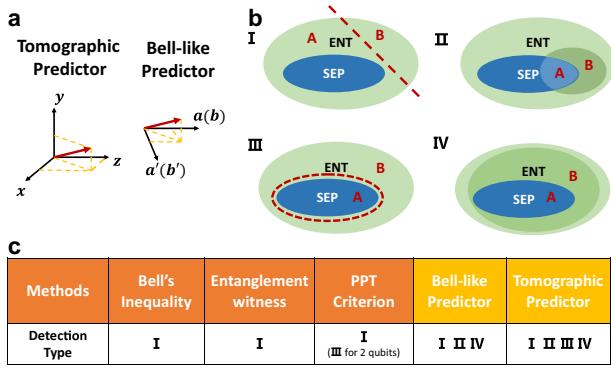
Fig. 1 Comparison between different methods. **a** Measurement angles for every qubit. For n-qubit system, the process of reading an unknown quantum state by standard quantum state tomography or some entanglement witness (e.g., the method depicted in Fig. 3b) requires to measure by three angles (σ_x, σ_y, σ_z) for every qubit. However, the detection of quantum entanglement by two-setting Bell's inequalities only requires two operators for each observer. **b, c** Various state classification tasks. Traditional tools aim to detect partial entangled states. The task belongs to type I. Our predictors aim to identify two (**a** and **b** in the figure) or more specific classes of quantum states. Note that 1 and 4 are intrinsically the same. In other words, if our training data cover all of separable states, the perfectly trained classifier can be regarded as entanglement witness

inputs of tomographic predictors are the Cartesian product of two sets of Pauli operators, $\{I, \sigma_x, \sigma_y, \sigma_z\}$, which contains a total of 15 non-trivial combinations. On the other hand, the CHSH operator in Eq. (7) can be regarded as an example of using the Bell-like predictors. There are various forms of Bell-like predictors in our paper. Only two rather than three local random operators for each qubit are used to constitute the inputs for all of Bell-like predictors, therefore their measurement resources is assumed to be smaller than the tomographic predictors for the same system.

To elaborate further, we construct a linear Bell-like predictor by generalizing the CHSH operator as (see Eq. (3) for notations):

$$\Pi_{ml} \equiv w_1 a_0 b_0 + w_2 a_0 b_0' + w_3 a_0' b_0 + w_4 a_0' b_0' + w_0, \qquad (10)$$

where the coefficients (or weights) $\{w_0, w_1, w_2, w_3, w_4\}$ are determined by the method of machine learning, through minimizing the error of detecting quantum entanglement of a given quantum ensemble.

Here the measurement angles $\{a_0, a_0', b_0, b_0'\}$ are taken to be the same as those given in Π_{CHSH} defined in Eq. (7). We denote the resulting Bell-like predictor as $CHSH_{ml}$. For a given quantum state, the set of measurement outcomes

$$\{\langle a_0 b_0 \rangle, \langle a_0 b_0' \rangle, \langle a_0' b_0 \rangle, \langle a_0' b_0' \rangle\}, \qquad (11)$$

are taken as the input of machine learning program. These elements are called "features" in the machine-learning literature. Normally, the number of elements in this set should be much smaller than the dimension of the quantum state.

In fact, the method of machine learning allows us to construct more general Bell-like predictors, given the same number of features. The key element of them is the inclusion of an extra hidden layer of neurons (Fig. 2c), compared with the linear predictor $CHSH_{ml}$. Moreover, each link between a pair of neurons is associated with a weight to be optimized in the learning phase.

Specifically, here we consider a class of non-linear predictors denoted by

$$Bell_{ml}(n, n_f, n_e), \qquad (12)$$

where n labels the number of qubits in the quantum state, n_f labels the number of features, and n_e labels the number of

neurons in the hidden layer of the neuron network. Apart from the extra neurons in the hidden layer, the measurement angles $\{a, a', b, b'\}$ in the corresponding feature list are taken randomly. For $n = 2$, the list is $\{ab, ab', a'b, a'b'\}$. See Table 1 for more details about the comparison of $Bell_{ml}$, and CHSH inequality.

In this work, all random measurement angles are obtained by $U\sigma_z U^\dagger$, where U is implemented by directly calling the function RandomUnitary.[30] Numerically, we found that they are uniformly distributed on the Bloch sphere. Furthermore, the mismatch rates are not sensitive to the choice of the measurement angles, when the number of neurons in the hidden layer is sufficiently large. The features of Bell-like predictors are obtained for a single set of random measurement angles in our work.

Labelling quantum states
As the first "test run" of our machine learning method, we focus on the following family of quantum states:

$$\rho_{\theta,\phi} = p|\psi_{\theta,\phi}\rangle\langle\psi_{\theta,\phi}| + (1-p)I/4, \qquad (13)$$

where $|\psi_{\theta,\phi}\rangle$ is defined in Eq. (8), $0 \leq p \leq 1$. For a pair of qubits, the entanglement between them can be determined by checking the PPT (positive partial transpose) criterion:[31] let $\rho_{\theta,\phi}^{T_B}$ be the matrix obtained by taking partial transpose of $\rho_{\theta,\phi}$ in the second qubit. The state is entangled if and only if the smallest eigenvalue of the matrix $\rho_{\theta,\phi}^{T_B}$ is negative. For n-qubit general system, in order to apply the PPT criterion, the full density matrix must be available, in order to obtain the minimum eigenvalue of the partial-transposed density matrix. However, state tomography requires an exponential number ($4^n - 1$) of measurements.

For our case, the minimal eigenvalue (the absolute of this value for entangled states is named as negativity[28]), can be obtained analytically (see Supplementary Materials for a derivation),

$$\lambda_{min}\left(\rho_{\theta,\phi}^{T_B}\right) = (1-p)/4 - p\cos(\theta/2)\sin(\theta/2). \qquad (14)$$

For each quantum state in the training set, we first evaluate the value of $\lambda_{min}\left(\rho_{\theta,\phi}^{T_B}\right)$, in order to create a label for it. In Fig. 3a, we depict the portion of separable states in the colored area of a Bloch sphere.

Testing phase of linear predictor
After the predictor is well-trained (see "Methods"), we test the performance by creating a new set of quantum ensemble that is distinct from the data set employed for training. Here the testing data comes from an ensemble of quantum states $\rho_{\theta,\phi}$ with a uniform distribution of p, θ and ϕ. Note that from Eq. (14), the entanglement of $\rho_{\theta,\phi}$ depends on the values of p and θ but not ϕ. However, the same set of features of the new density matrices are provided as the input; the values of p and θ are not directly provided in the testing phase, but they are used to evaluate the performance of the predictors.

We quantify the performance of the $CHSH_{ml}$ predictor as follows: for given values of p and θ, the mismatch rate $R_{mm}(p, \theta)$ is defined by the probability that the function outputs a different label from the PPT criterion, averaged over uniform distribution of the angle ϕ, i.e.,

$$R_{mm}(p, \theta) \equiv \Pr(1_{ML}|0_{PPT}) + \Pr(0_{ML}|1_{PPT}), \qquad (15)$$

where $x_{ML} \in \{0, 1\}$ labels the output of the machine learning predictor; 1_{ML} (0_{ML}) means separable (entangled), and similarly for x_{PPT}. Of course, the match rate can be defined in a similar way (i.e., $1 - R_{mm}$).

First, we only trained and tested with data on fixed $\phi = 0$, $CHSH_{ml}$ preforms satisfactory for any value of θ. The form of this
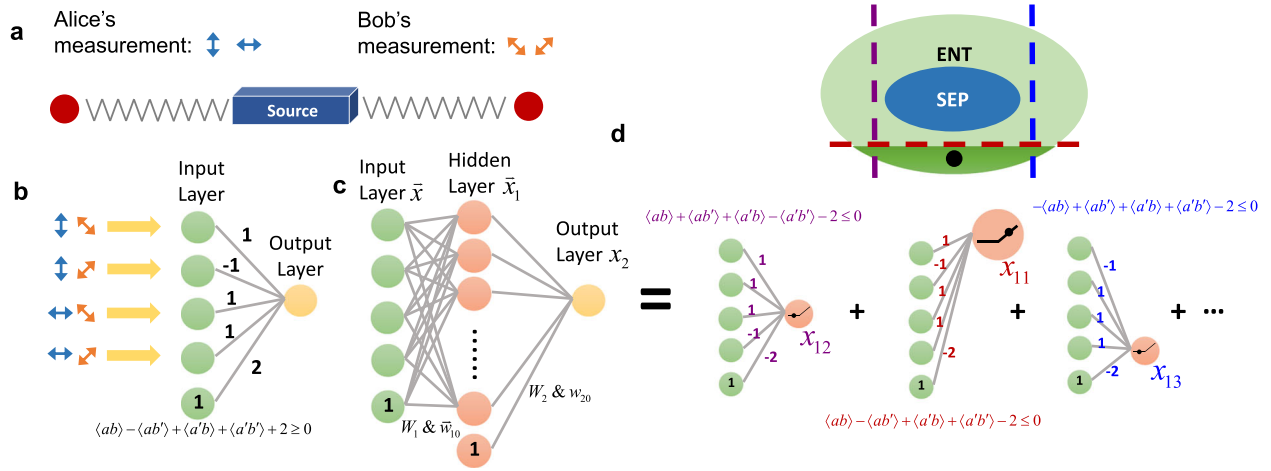
Fig. 2 CHSH inequality and machine learning. **a** Typical setup for obtaining CHSH inequality. **b** Encoding CHSH inequality to the simplest network (called Rosenblatt's perceptron[24]). **c** Artificial Neural Network with hidden Layers. The objective of machine learning is optimizing $\sigma_S(W_2((W_1\vec{x} + \vec{w}_{10})) + w_{20})$ where σ_{RL} is ReLu function and $\sigma_S(z) = 1/(1 + e^{-z})$ is sigmoid function. **d** The neural network with a hidden layer and ReLu function can be regarded as encoding multiple CHSH (witness) inequalities simultaneously in a network

becomes: $-14\langle \boldsymbol{a_0 b_0}\rangle - 28\langle \boldsymbol{a_0' b_0'}\rangle + 10$. Here we keep the trained parameters in integer values.

Then, we trained our model again with different values of ϕ. Through a linear optimization process, the form of this is numerically found to be: $0.521\langle \boldsymbol{a_0 b_0}\rangle - 0.603\langle \boldsymbol{a_0 b_0'}\rangle - 0.025\langle \boldsymbol{a_0' b_0}\rangle + 0.016\langle \boldsymbol{a_0' b_0'}\rangle + 0.373$. As shown in Fig. 3c, d, can classify the data as entangled (or separable) state if $p > 1/2$ ($p \le 1/2$). Although it is not perfect, the predictor yields a better performance than the standard CHSH for most of the testing states.

To be specific, let us focus on the state of $\theta = \pi/2$ (Fig. 3c). The numerical data indicates that both CHSH and CHSH$_{ml}$ can identify the regime where $\lambda_{min}\left(\rho_{\theta,\phi}^{T_B}\right) > 0$ as separable. Beyond that region, CHSH results in a 100% mismatch rate, but CHSH$_{ml}$ can reduce the mismatch rate as p increases. Therefore, the performance of CHSH$_{ml}$ is significantly better than that of CHSH in identifying quantum entanglement. The reason for CHSH to produce 100% mismatch rate is explained after Eq. (6): there exist entangled states not violating the CHSH inequality for any choice of ϕ. Besides, the limitations of CHSH$_{ml}$ are due to the following facts:

1. The features taken directly from the original CHSH inequality do not include any information about σ_y.
2. As an entanglement witness, any Bell-like inequality is not sufficient to characterize the boundary between entangled and separable states on our "test run" states.

Next, we shall see that the performance of machine learning can be significantly increased, if we choose to make the measurement angle random and add a hidden layer, i.e., Bell-like predictor.

Encoding Bell's inequalities in a neural network

The key idea of the non-linear model can be regarded as a transformation from a group of Bell's inequalities or entanglement witnesses. For traditional methods, any quantum state must be entangled if it violates at least one Bell's inequality. Here different inequalities have different weights. For example, Eq. (3) can be regarded as two different inequalities,

$$-\langle \boldsymbol{ab}\rangle + \langle \boldsymbol{ab'}\rangle - \langle \boldsymbol{a'b}\rangle - \langle \boldsymbol{a'b'}\rangle - 20, \tag{16}$$

and

$$\langle \boldsymbol{ab}\rangle - \langle \boldsymbol{ab'}\rangle + \langle \boldsymbol{a'b}\rangle + \langle \boldsymbol{a'b'}\rangle - 20. \tag{17}$$

By swapping $\boldsymbol{a}(\boldsymbol{b})$ and $\boldsymbol{a'}(\boldsymbol{b'})$ of Eq. (17), we obtain some variants of the CHSH inequalities

$$\langle \boldsymbol{ab}\rangle + \langle \boldsymbol{ab'}\rangle + \langle \boldsymbol{a'b}\rangle - \langle \boldsymbol{a'b'}\rangle - 20, \tag{18}$$

$$-\langle \boldsymbol{ab}\rangle + \langle \boldsymbol{ab'}\rangle + \langle \boldsymbol{a'b}\rangle + \langle \boldsymbol{a'b'}\rangle - 20. \tag{19}$$

When applied separately, these new CHSH inequalities potentially detect different entangled states.

In fact, all of these variants can be encoded into one neural network model with a hidden layer (see Fig. 2c, d). Here the connections between each hidden neuron (except bias) and input layer corresponds to one Bell's inequality. Specifically, we apply ReLu[32] function

$$(z)_{RL} = \max(z, 0), \tag{20}$$

on every hidden neuron to ensure that the output is always 0 if the state do not violate any Bell's inequality. Otherwise, the entanglement state can be detected and quantified by the hidden neuron.

Furthermore, for any entangled state, there *exists* at least one entanglement witness detecting it (known as the completeness of witnesses[28]). Because all of these "witness inequalities" can be encoded into one neural network, our tomographic predictor represents a generic entangled-separable state classifier; there *exists* a set of weights in the ANN neural network for distinguishing *any* finite set of entangled states from all separable states. The size of the hidden layer is not larger than that of the entangled states. More details about the formal argument, together with the definition of witness and ANN formulas, are placed in the "Methods" section.

Testing phase of Bell-like predictor

The result of three predictors, Bell$_{ml}$(2, 4, x) (i.e., for 2 qubits, 4 features, and x neurons in the hidden layer, here $x = 0, 20, 150$) are shown in Fig. 3e, f. The overall performance in terms of mismatch rates are significantly improved, compared with the CHSH$_{ml}$ predictor. Furthermore, inclusion of a hidden layer can significantly mitigate the problem of CHSH$_{ml}$ near $\theta = 0$. Note that the state $|\psi_{\theta,\phi}\rangle$ with $\theta = 0$ reduces to only one state $|00\rangle$ for any choice of ϕ, thus the mismatch rate becomes 100% whenever the predictor made a mistake.

We note that such a problem in CHSH$_{ml}$ also exists in the Bell$_{ml}$ predictor without hidden layer. However, the problem goes away whenever hidden layers are included. Numerically, we find that

Table 1. CHSH inequality versus machine learning predictors for two qubits

Type	Features	Weights	Form
CHSH	fixed $\{a_0 b_0,\, a_0 b_0',\, a_0' b_0,\, a_0' b_0'\}$[a]	4 Fixed values[b]	Linear, no hidden layer
CHSH$_{ml}$	fixed $\{a_0 b_0,\, a_0 b_0',\, a_0' b_0,\, a_0' b_0'\}$[c]	4 Optimized values	Linear, no hidden layer
Bell$_{ml}$	variable $\{ab,\, ab',\, a'b,\, a'b'\}$[d]	many optimized values	Non-linear, with hidden layer

[a]$a_0 = \sigma_z$, $a_0' = \sigma_x$, $b_0 = (\sigma_x - \sigma_z)/\sqrt{2}$, $b_0' = (\sigma_x + \sigma_z)/\sqrt{2}$
[b]$w_0 = \pm 2$, $w_1 = 1$, $w_2 = -1$, $w_3 = 1$, $w_4 = 1$
[c]For "test run" states, $\langle a_0 b_0 \rangle = \langle a_0 b_0' \rangle$ and $\langle a_0' b_0 \rangle = \langle a_0' b_0' \rangle$, the features of CHSHml contains redundant information. If only trained with $\langle a_0 b_0' \rangle$ and $\langle a_0' b_0' \rangle$, the predictor also works well
[d]a, b, a', and b' are randomly generated and uniformly distributed on the Bloch sphere

the results with a total of 150 neurons in the hidden layer do not significantly out perform the results with only 20 neurons.

Classifying general two-qubit states

In the previous section, we have studied the ability of machine-learning predictors in identifying the entanglement of quantum states of the form indicated in Eq. (13), which belongs to type II (Fig. 1b) problem. Although Bell-like predictors perform better than the CHSH inequality for some quantum states, a more interesting question is, can we construct a universal function that accepts only partial information about the quantum state but, at the same time, can detect all entangled states by machine learning (i.e., type III)? A negative result suggests that such a classifier may not exist.[33] Therefore, in order to train a universal entanglement classifier, tomographic predictor should be considered.

Tomographic predictor is universal for such task, in the sense that one can classify all separable and all entangled states with an *infinite* number of neurons in the hidden layer. Therefore, an important and interesting question is, are standard machine-learning algorithms capable of training a *finite* neural network for such a task at a high accuracy?

In this section, we verify that the answer is true. For the case of two qubits, we can still rely on the PPT criterion to provide labels for our training set. For this part, we generate a new training set of 2-qubit mixed states randomly and label them by the PPT criterion in the same way as the previous section. The ensembles ρ are prepared by first generating a set of random matrices σ, where the real and imaginary parts of the elements $\sigma_{ij} = a_{ij} + i b_{ij}$ are generated by a Gaussian distribution with a zero mean and unit variance. The resulting density matrix is obtained by

$$\rho_{rand} = \sigma \sigma^{\dagger} / (\sigma \sigma^{\dagger}), \qquad (21)$$

which is implemented by using the code of RandomDensityMatrix.[30]

The performance of our machine-learning predictors heavily depends not only on the training set but also the distribution of the testing data. We find that many data points are localized near the boundary between entangled and separable states, which represents a challenge for us; machine learning performs not so well near the marginal cases.

The distribution of λ_{min} in our data set is given in Fig. 4a. We can see that the majority of states are weakly entangled, which imposes the challenge for our machine-learning predictor. The population of entangled states in our data set is about 75%. To avoid a bias in our training set, the fraction of entangled states is about the same as separable states, as shown in the red area in Fig. 4a. However, all new data are used in the testing stage. The mismatch rates of both separable (blue) and entangled (green) data are listed in Fig. 4b individually, showing the increase in the performance of the tomographic predictor as the number of hidden neuron increases. Furthermore, the mismatch rate of states with different λ_{min} is depicted in Fig. 4c; the network becomes

more reliable when larger size of hidden layer units are available. Small fraction of error occurs near the boundary of entangled and separable states. The mismatch rate decreases to about 0.5% with 3200 hidden units, if λ_{min} of test data distributes uniformly between -0.38 and 0.14, rather than near the boundary of separable-entangled hyperplane.

Note that our tomographic predictor are trained from scratch without exploiting any prior information (the weights are initialized randomly by standard API), but the performance of trained hidden layer are indeed similar to entanglement witness. The results are depicted in Fig. 4d–f and the details shall be discussed in the "Methods" section.

The results in this section demonstrate the capability of our tomographic predictor to detect unknown entangled states. It paves a way to the development of a generic tool for entanglement detection in other intricate, for example, 3 by 3 (2-qutrit) and 3-qubit systems. If the qubit number $n \geq 3$, all of traditional methods such as PPT and entanglement witness are only expected to detect partial entangled states (type I in Fig. 1b). Even if a complete characterization of a quantum state is given, the task of determining whether it is entangled or not (type III) is time-consuming with numerical tools, such as semidefinite programming. If the state is labelled by such numerical tool, our machine learning methods would potentially reduce the time significantly for predicting the class of new states.

Machine learning for identifying bound entangled states

The entanglement structure of a three-qubit system is significantly more complicated than two-qubit systems. As seen in Fig. 5b, it can be classified into several types of entanglement classes.[28] In particular, a three-qubit quantum state is called "biseparable", if two of the qubits are entangled with each other but not with the third one. The corresponding density matrices are denoted as $\{\rho_{A|BC}, \rho_{B|AC}, \rho_{C|AB}\}$ and their convex combination, i.e., $\lambda_1 \rho_{A|BC} + \lambda_2 \rho_{B|AC} + \lambda_3 \rho_{C|AB}$ for $0 \leq \lambda_1, \lambda_2, \lambda_3 \leq 1$ and $\lambda_1 + \lambda_2 + \lambda_3 = 1$. Of course, these sets of states include fully-separable states as a special case. A system is called fully-entangled[28] if it is neither biseparable nor fully-separable.

There are two types of typical fully-entangled states.

1. Any 2 qubits are entangled with each other. For these states, any partial reduced matrix, $x(\rho_{ABC})$ for $X \in \{A, B, C\}$, should be detected as entangled states by the PPT criterion. A typical example is the W state, i.e. $(|100\rangle + |010\rangle + |001\rangle)/\sqrt{3}$.
2. Any 2 qubits are separable. If given all of information about density matrix, some of them can be identified as entangled states by PPT criterion, for example, Greenberger-Horne-Zeilinger (GHZ) state $(|000\rangle + |111\rangle)/\sqrt{2}$, while others are not and called as bound entangled states.

Numerically, we found that almost all of ensembles generated by Eq. (21) are entangled states and can be detected by PPT.

We therefore focus on another task: training a tomographic or Bell-like predictor to identify bound entangled states and fully
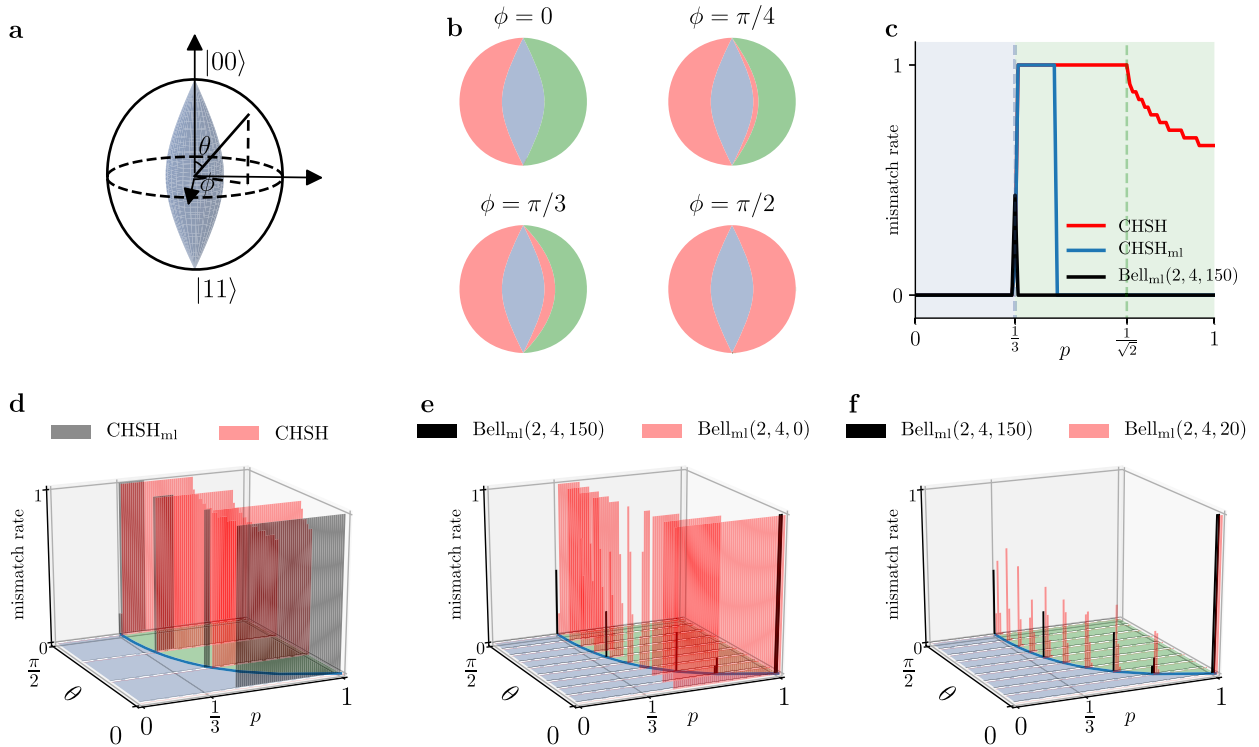
Fig. 3 Results of the first "test run" by machine learning. **a** The "shape" of quantum states illustrated by Bloch sphere; the blue area represents separable states. **b** Limitation of entanglement witness detection. The entangled states detected by single witness $\mathcal{W}$ depends on at least three angles[28] and the state phase ϕ. As an example, the entangled states which lie in green area can be detected, while that in red area can not. See Supplementary Materials for more details. **c, d** The optimization of original CHSH inequality by tuning W and w_0. For states $\rho_{\theta,\phi}$ (Eq. (13)) with fixed θ, p but different angle ϕ, the height of vertical axis presents the mismatch (i.e., error) rate. Here $p \in [0, 1]$ are uniformly divided into 100 parts and $\phi \in [-\pi, \pi)$ into 60 parts, same with (**d–f**). **c** is the cross section of (**d–f**) with $\theta = \pi/2$, which illustrates the optimization of maximum entangled states. **e, f** Mismatch rate of entanglement detection by Bell-like predictor with different hidden layers on 2-qubit system

separable states (Fig. 5a); neither of them can be identified as "entangled states" through PPT criterion.

In 1998, Bennett[34] showed that in a group of four product states in a 3-qubit system, the complementary counterpoints are bound entangled states. Theses product states are called unextendible product basis (UPB) and denoted by $\{|v_i\rangle\}$, $i = 1, 2, 3, 4$, where $|v_1\rangle = |000\rangle$, $|v_2\rangle = \frac{1}{2}|1\rangle(|1\rangle - |0\rangle)(|1\rangle + |0\rangle)$, $|v_3\rangle = \frac{1}{2}(|1\rangle + |0\rangle)|1\rangle(|1\rangle - |0\rangle)$, $|v_4\rangle = \frac{1}{2}(|1\rangle - |0\rangle)(|1\rangle + |0\rangle)|1\rangle$.[30] The normalized form of

$$\rho_{\text{tile}} = I - \sum_i |v_i\rangle\langle v_i|, \tag{22}$$

is a bound entangled state.

To generate sufficiently many bound entangled states for training, we applied operations called stochastic local operations assisted by classical communication (SLOCC).[35] Specifically, three independent *random* matrices $\sigma = \sigma_A \otimes \sigma_B \otimes \sigma_C$ are applied on each qubit of ρ_{tile}, i.e.,

$$\rho = \frac{\sigma \rho_{\text{tile}} \sigma^\dagger}{\text{Tr}(\sigma \rho_{\text{tile}} \sigma^\dagger)}. \tag{23}$$

Here the elements of each matrix $\sigma_{A,B,C} = a_{ij} + ib_{ij}$ were discussed in the previous section. Numerically we found that all of the generated random matrices are invertible. Moreover, fully-separable states are obtained by the sum of random product ensembles according to its definition in Eq. (2).

In our implementation, similar to our previous construction of Bell-like predictors based on Bell's inequalities, here we consider the Mermin inequality[36] and Svetlichny inequality[37] as the starting points. For three-qubit systems, the Mermin inequality is of the form

$$|\langle abc \rangle - \langle ab'c' \rangle - \langle a'bc' \rangle - \langle a'b'c \rangle| \leq 2. \tag{24}$$

The Svetlichny inequality (essentially a double Mermin inequality) is of the following form

$$|\langle abc' \rangle + \langle ab'c \rangle + \langle a'bc \rangle - \langle a'b'c' \rangle + \langle a'b'c \rangle + \langle a'bc' \rangle + \langle ab'c' \rangle - \langle abc \rangle| \leq 4. \tag{25}$$

The Mermin inequality and the Svetlichny inequality are the multipartite counterpart of Bell's inequalities. Therefore, one can also employ these inequalities for detecting multipartite entanglement. In a similar way, we can also apply the machine learning method to boost the efficiency.

In our machine learning method, we adopted the elements of Mermin inequality as input (four features) to train our Bell-like predictor $\text{Bell}_{\text{ml}}(3, 4, x)$, and similarly, for the Svetlichny inequality, we constructed $\text{Bell}_{\text{ml}}(3, 8, x)$.

The mismatch rate of the machine learning method is shown in Fig. 5c. It is indicated that if we just use the same number of features as in Mermin $\text{Bell}_{\text{ml}}(3, 4, x)$ and Svetlichny $\text{Bell}_{\text{ml}}(3, 8, x)$ inequalities, the performance is not satisfactory. The mismatch rate cannot get much improved by increasing the number of neurons in the hidden layer. However, the performance of machine learning method can get significantly improved by putting three groups of features from CHSH inequalities for every pair of qubits, which gives a new $\text{Bell}_{\text{ml}}(3, 12, x)$ predictor. As a benchmark, the mismatch rate of tomographic predictor can be decreased to nearly 0%. If given more information, Bell-like predictor performs almost same (about 1%) as the tomographic predictor. For example, for the Bell-like predictor, $\text{Bell}_{\text{ml}}(3, 26, x)$, the 26 features are generated in the following way: assume there
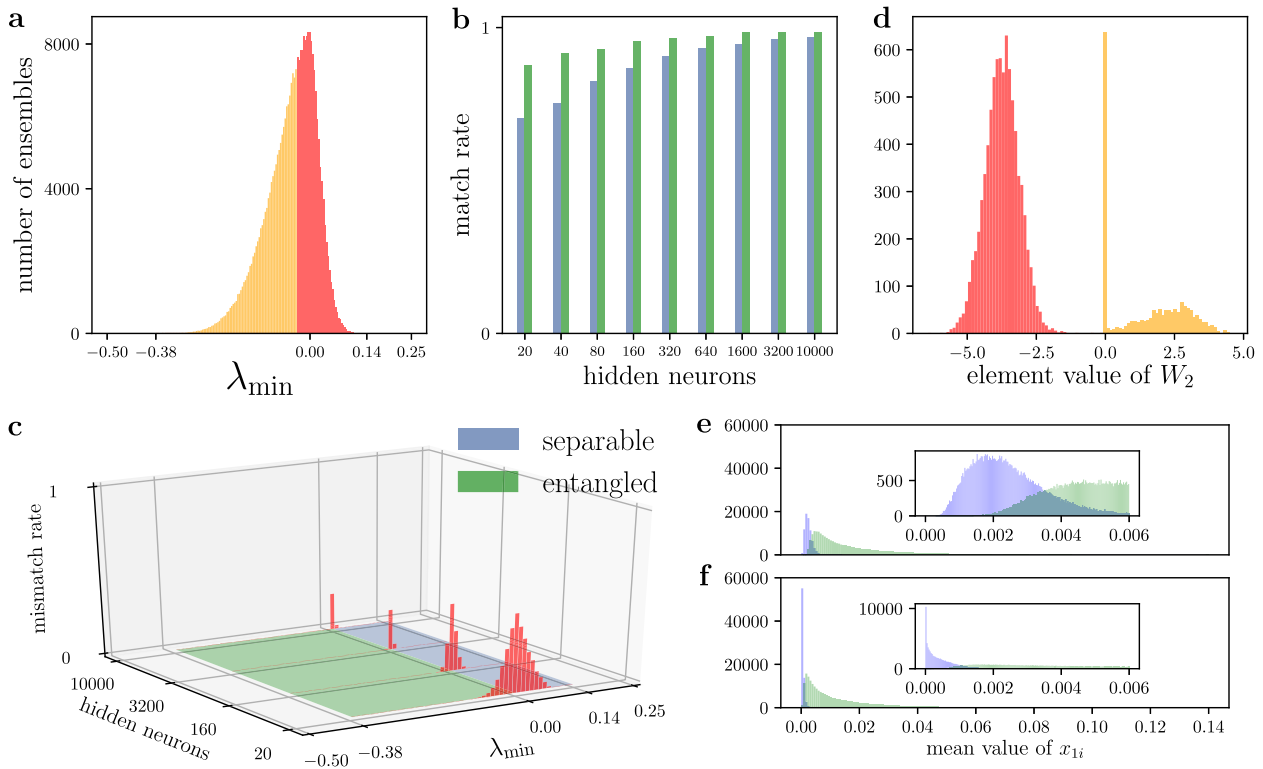
Fig. 4 Tomographic predictor on all 2-qubit states. **a** Histogram of $\lambda_{\min}\left(\rho_{AB}^{T_B}\right)$. In our numerical results, 3,000,000 data are generated but only the data on boundary (red area) are used for training. Then 300,000 new data are generated for testing. The vertical axis here refers to test data. **b** Match rate of 300,000 test states predicted by tomographic predictors on the general 2-qubit ensembles. **c** Mismatch rate (test data) of fixed $\lambda_{\min}$ with different length of hidden layer units. **d** Histogram of the trained w_{2i}. Here $W_2 = [w_{21}, w_{22}, \cdots w_{2n_e}]$ with $n_e = 10,000$. **e** Distribution of $\frac{1}{n_e}\sum_{i=1}^{n_e} x_{1i}$ for 300,000 test data. **f** Same with (**e**) but excluding x_{1i} if $w_{2i} \geq -0.01$ (the orange area in (**d**))

are three parties and each party performs a measurement on a qubit locally in two different angles labelled by $\hat{n}^i$, $\hat{n}'^i$ ($i = 1, 2, 3$). Then, a feature is obtained by the joint expectation value $\langle O^1 O^2 O^3 \rangle$, where $O \in \{\hat{n}, \hat{n}', I\}$. Note that $I^1 I^2 I^3$ is excluded, since $\langle I^1 I^2 I^3 \rangle = 1$ for any quantum state. The feature number decreases from $4^n - 1 = 63$ for tomographic predictor, to $3^n - 1 = 26$ for Bell-like predictor with a similar performance.

DISCUSSION

In this work, we have applied a method of machine learning, known as artificial neuron networks (ANN), to solve problems of entanglement-separable classification in quantum information science. We have achieved several results, including

1. Optimization of CHSH inequality or Bell-type inequalities. Our machine-learning architecture can yield a much better performance for a class of testing states.
2. Exploring the challenges for constructing a universal entanglement detector for two-qubit systems.
3. A novel physical interpretation of network-based model from the perspective of quantum information. As an entanglement-separable classifier, we presented an argument showing that the tomographic model is universal in the large-N limit. We have numerically studied the trained weights and found that their performance is very similar to entanglement witnesses. The result is consistent with our interpretation that each witness can be encoded in the hidden layer of network. The details are documented in the "Methods" section.
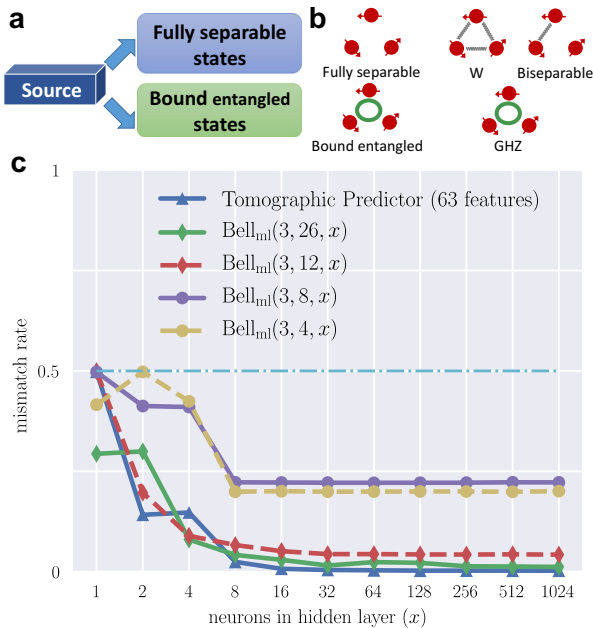
Fig. 5 Biseparable and bound entangled states distinguished by Bell-like and tomographic predictors. **a** The generation process of 3-qubit states. 200,000 bound entangled and 200,000 fully separable data are generated individually. Ninety percent of them are used for training and others for testing. **b** Different quantum states in 3-qubit system. **c** Five types of predictors applied on the ensembles

4. Construction of both tomographic and Bell-like predictors to classify quantum states that cannot be detected by PPT criterion.

Overall, we found that machine-learning can produce reliable results, given a proper training set of data. The performance of machine learning becomes worse whenever the majority of quantum states in the training set lies around the boundary between two classes (e.g., entangled and separable) of quantum states.

One may ask what if we can apply our methods to general quantum systems with unknown entanglement or large qubit number. The answer depends on the task we aim at.

1. For n-qubit (few body) states where the entanglement is completely unknown. Tomography is proved to be necessary for universal entanglement detection.[33] Therefore any algorithm of detecting unknown entangled states is expected to depend on the measurement outcomes of $4^n - 1$ observables. Although we have argued that there exist the tomographic predictor's weights for detecting *any* finite set of entangled states, the numerical validity of tomographic predictor is still an open question due to the difficulties in making correct labels and generating appropriate quantum states.

2. For n-qubit (few body) system where the entanglement is partially known. PPT criterion fails for detecting the so-called bound entangled states. We find that Bell-like predictors with $3^n - 1$ observables has good performance for identifying bound entangled and fully separable states. Next, in the Supplementary Materials, as an extension, we consider multiple-state classification involving the system of three qubits. We construct a quantum state classifier that can identify four types of states, including three types of entangled (biseparable) and one type of fully-separable states. Furthermore, we consider the ensembles of four-qubit system. We train and analyse the performance of state classifiers in terms of three groups of quantum states, including entangled, separable ones and the states without correct label. The performances of Bell-like predictors using $3^n - 1$ features in all of these systems are satisfactory, although the form of entangled states are different from each other.

3. For n-qubit (many body) states where the form of entanglement is partially known. In the Supplementary Materials, we have studied identifying n-qubit GHZ-type states by non-tomographic predictors. Any $n - 1$ particles of GHZ state constitute a fully separable system. Therefore, in order to distinguish GHZ-type states from fully separable states, the application of PPT criterion requires $4^n - 1$ measurement outcomes (features) and a diagonalization of an exponentially-large matrix. We found that our ANN architecture are capable to identify fully separable and various GHZ-type states using only $2n$ random features.

In our approach, the goal of machine learning method is to "learn" the labels we assign to a quantum state. This is the reason we use "match/mismatch rate", instead of using "error". In other words, if we label some states with mistakes, the machine learning methods may learn the mistakes as well, unless most of the same states are labelled correctly. Without additional instructions, all of data shown in our figures are test data. And all of error/accuracy (mismatch/match rate) refers to test error/accuracy. Overall, for scaling up this method for detecting higher-dimensional quantum entanglement, the major challenge is related to a lack of reliable method for labeling the entanglement. One possible direction to further explore is through labeling entanglement with semi-definite programming (SDP), which is a highly time-consuming process. Our machine learning method can potentially speedup the state-classification process by learning the SDP labeling.

In general, our results imply that machine learning is particularly useful for problems where the process of labelling a quantum state is resource consuming. A significant contribution of this work is that it reveals the relationship between a widely used machine learning architecture called ANN and the theory of entanglement witnesses from both analytical and numerical perspectives.

METHODS

General background on quantum entanglement

Entanglement is a key feature of quantum mechanics, where the correlation between pairs or groups of particles cannot be described within a local realistic classical model. In quantum information theory, entanglement is regarded as an important resource to achieving tasks, such as quantum teleportation, computation, and cryptography. However, given a quantum state, the problem of determining if it is entangled or not is a computationally-hard, this question is particularly important in quantum experiments. Currently, methods of entanglement detection has been developed for specific scenarios. The most popular ones includes positive partial transpose (PPT) criteria[31] and entanglement witnesses.[28]

For a pair of qubits, PPT is both sufficient and necessary for entanglement detection.[38] However, PPT is a necessary but not sufficient condition for multi-qubit systems. In addition, it requires the knowledge of the whole density matrix. Experimentally, it means one needs to perform quantum state tomography, which is resource consuming for multiple-qubit systems.

Moreover, entanglement witnesses represent a different approach for entanglement detection. Given an observable $\mathcal{W}$, where $\text{Tr}(\mathcal{W}\rho) \geq 0$ for all separable states. If $\text{Tr}(\mathcal{W}\rho) < 0$ for (at least) one entangled ρ, then we say $\mathcal{W}$ detects ρ.[28] Here the trace $\text{Tr}(\mathcal{W}\rho) = \langle \mathcal{W} \rangle$ represents the measurement result of ρ with $\mathcal{W}$. Although any entangled states can be detected by at least one witness,[28] there is no efficient way to find it out. In other words, it is possible that there are entangled states not detected by a given witness, i.e., $\text{Tr}(\mathcal{W}\rho) \geq 0$ for an entangled state.

On the other hand, quantum entanglement is necessary for a violation of Bell's inequalities,[25] which has been confirmed in numerous experiments.[39–41] In principle, Bell's inequality can be employed for detecting quantum entanglement; it can witness some entangled states. It is an attractive direction, as only partial information is needed from the quantum state. However, for normal Bell's inequalities, only small part of the entangled states can be detected; a situation similar to entanglement witness. Motivated by this problem, one of our objective is to construct a quantum-state classifier for entanglement detection through optimizing Bell's inequalities.

Overview of ANN with single hidden layer

Consider the scenarios where quantum states are distributed to different parties through a noise channel characterized by some unknown parameter. The parties are given the opportunity to test the channel through a set of testing states, which corresponds to the training phase of machine learning. At the end, the parties are given an non-linear function optimized for the purpose of state classification, where only partial information is required for testing new quantum states beyond the training set.

Our non-linear quantum-state classifier is constructed by a technique in machine learning known as artificial multilayer perceptron,[24] which is a network composed of several layers, where information flows from input layer, through hidden layer, and finally to the output layer.

The input layer contains the information about the quantum state, where the expectation value of certain observables are taken as the elements of a vector $\vec{x}$. The hidden layer contains another vector $\vec{x}_1$, which is constructed through the relation,

$$\vec{x}_1 = \sigma_{RL}(W_1 \vec{x} + \vec{w}_{10}). \tag{26}$$

Here W_1, $\vec{w}_{10}$ are initialized uniformly and optimized through the learning process. And ReLu function[32] defined by $\sigma_{RL}\left([z_1, z_2, \cdots, z_{n_e}]^T\right) = [\max(z_1, 0), \max(z_2, 0), \cdots, max(z_{n_e}, 0)]^T$, is a non-linear function for every neuron. Finally, the neuron(s) in the output layer contains the probabilities for the input state to belong to a specific class. For example, for a binary-state classification, where only one neuron is needed, the output y contains the probability for the input state may be identified as entangled

or separable. Here

$$y = \sigma_S(W_2 \vec{x}_1 + w_{20}), \qquad (27)$$

where $\sigma_S(z)$ is sigmoid function

$$\sigma_S(z) = 1/(1 + e^{-z}). \qquad (28)$$

For non-linear predictors, both W_2 and w_{20} are parameters to be trained. For linear predictor (CHSH_{ml}), $W_2 = 1$ and $w_{20} = 0$.

Universality of tomographic predictor

In this section, we shall argue that the tomographic predictor is generic to classify any set of entangled and separable states. More precisely, all of separable states and a finite set of any entangled states can be distinguished by our model. The size of hidden layer is expected to be not larger than that of entangled states. The main ingredient is a theorem called *completeness of witnesses*:[28] *for any entangled state, there exists at least one entanglement witness detecting it.*

According to the definition of witness and the fact that different Hermitian matrices can be expanded as the sum of a finite set of fixed base observables, i.e., $\mathcal{W} = \sum_i w_i \hat{o}_i$. For n-qubit system, the set $\{\hat{o}_i\}$ has $4^n - 1$ elements. We consider the following "witness inequalities",

$$\sum_i w_i \langle \hat{o}_i \rangle < 0 \text{ for some entangled state},$$
$$\sum_i w_i \langle \hat{o}_i \rangle \geq 0 \text{ for all separable states.} \qquad (29)$$

For 2-qubit systems, the set is just the Cartesian product of two identical sets of Pauli operators, $\{I, \sigma_x, \sigma_y, \sigma_z\}$. In the section "Encoding Bell's inequalities in a neural network", we introduce how to encode different groups of $\{w_i\}$ on an ANN architecture with one hidden layer. (A minor issue is that here w_i should be replaced by $-w_i$ to be consistent with the main text). As seen in Eq. (26) and Fig. 2c, d, expanding $\vec{x}_1 = [x_{11}, x_{12}, \cdots x_{1n_e}]^T$, each hidden neuron x_{1i} encodes the detection result of corresponding witness. If the state does not violate this witness inequality, $x_{1i} = 0$, otherwise $x_{1i} > 0$. Therefore

$$\sum_i x_{1i} \geq 0 \text{ for all entangled states},$$
$$\sum_i x_{1i} = 0 \text{ for all separable states.} \qquad (30)$$

If each entangled state can be detected by at least one witness, "≥" can be turned to ">".

Define $W_2 = [w_{21}, w_{22}, \cdots w_{2n_e}]$. According to Eq. (27), if each $w_{2i} \leq 0$ for $i > 1$ and $w_{20} = 0$, we have

$$y \leq 0.5 \text{ for all entangled states},$$
$$y = 0.5 \text{ for all separable states.} \qquad (31)$$

"≤" can be turned to "<" if each entangled state can be detected by at least one witness. And the hidden layer can be compressed by removing the x_{1i} units if $w_{2i} = 0$.

Therefore, the weights of a "perfect" tomographic predictor should in principle exist; the remaining problem is, how to find out these weights? In this work, we aim at finding them out by machine-learning methods.

In the section "Classifying general two-qubit states", we trained such a predictor and found that with a sufficiently-large number of hidden neurons, tomographic predictor performance is satisfactory. As seen in Fig. 4d, most of the trained elements of W_2 are very close to 0. And Fig. 4e illustrates different performance of entangled and separable states. In most cases, the sum of x_i are larger than 0 for entangled states and close to 0 for separable states, which is consistent with our theory we discussed above. Figure 4f tells us if we only keep the negative part of W_2 and the corresponding hidden neurons, x_1 of separable states are more close to 0, which is also consistent with our arguement.

Training of the predictors

To investigate the performance of CHSH_{ml} and Bell-like predictors, which is essentially a linearly-optimized version of CHSH, and a non-linear predictor with machine learning (see Fig. 2b, c). First, we need to generate an initial set of quantum states, called training set. for "test run" states in our first model (Eq. (13)), the set of 200,000 states are generated by sampling a uniform distribution of θ and ϕ, but with a Gaussian distribution for p, with a mean value $1/(1 + 2 \sin\theta)$, which yields an ensemble of states in the boundary of separable and entangled hyperplane.

Specifically, for each time, we evaluated the four features, like $\{\langle a_0 b_0 \rangle, \langle a_0 b'_0 \rangle, \langle a'_0 b_0 \rangle, \langle a'_0 b'_0 \rangle\}$ in the CHSH_{ml} for a given state in the training set, putting them into a four-dimensional feature vector $\vec{x}$ in ANN. In fact, if we consider only one side of the inequality, the CHSH inequality is equivalent to

$$W_0 \vec{x} + w_0 \geq 0, \qquad (32)$$

where $W_0 = [1, -1, 1, 1]$ and $w_0 = 2$. In other words, CHSH inequality are violated iff the output value is negative. The problem of optimization of CHSH_{ml} is equivalent to the problem of finding an optimal set of matrix elements for W and w_0 (for non-linear predictors they are $W_1, W_2, \vec{w}_{10}, w_{20}$), through the given training set of quantum state.

We make use of a loss function constructed by the binary or cross entropy[42] to calculate the difference between predictor and the results based on the PPT criterion for many copies in the given quantum ensemble. The entire implementation of training ANN architecture depends on the neural networks API keras.[43] The optimizer we chose is RMSprop with default hyper-parameters. For example, the learning rate is 0.001. At the end, we obtained a vector W and w_0 that is optimized by the above process.

Shortly after our original manuscript was posed on arXiv, Lu et al.[44] reported that they independently combined machine learning and semidefinite programming to train their predictors as quantum state classifiers. Using all information without any prior knowledge, the error of their predictor is always around 10% on general 2-qubit system. However, our tomographic predictor performs below 2% on the same ensembles with 3000 hidden neurons.

Data availability

The codes that support the findings of this study are available in figshare with the identifier. https://doi.org/10.6084/m9.figshare.6231662.[45]

AUTHOR CONTRIBUTIONS

Y.C.M. and M.H.Y. designed the project and analyzed the results; Y.C.M. implemeted the machine learning code. Both authors wrote the paper.

ADDITIONAL INFORMATION

Supplementary information accompanies the paper on the *npj Quantum Information* website (https://doi.org/10.1038/s41534-018-0081-3).

Competing interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

REFERENCES

1. Lloyd, S., Mohseni, M. & Rebentrost, P. Quantum principal component analysis. *Nat. Phys.* **10**, 631–633 (2014).
2. Ciliberto, C. et al. Quantum machine learning: a classical perspective. *Proc. R. Soc. A* **474**, 20170551 (2018).
3. Bishop, C. M. *Pattern recognition and machine learning* New York: Springer-Verlag (2006)..
4. Carleo, G. & Troyer, M. Solving the quantum many-body problem with artificial neural networks. *Science* **355**, 602–606 (2017).
5. Huang, L. & Wang, L. Accelerated monte carlo simulations with restricted boltzmann machines. *Phys. Rev. B* **95**, 035105 (2017).
6. Carrasquilla, J., & Melko, R. G. Machine learning phases of matter. *Nature Physics* **13**, 431 (2017).
7. Schoenholz, S. S., Cubuk, E. D., Sussman, D. M., Kaxiras, E., & Liu, A. J. A structural approach to relaxation in glassy liquids. *Nat. Phys* **12**, 469 (2016).
8. van Nieuwenburg, E. P., Liu, Y.-H. & Huber, S. D. Learning phase transitions by confusion. *Nat. Phys.* **13**, 435–439 (2017).

9. Deng, D.-L., Li, X. & Sarma, S. D. Quantum entanglement in neural network states. *Phys. Rev. X* **7**, 021021 (2017).

10. Torlai, G. et al. Many-body quantum state tomography with neural networks. *arXiv preprint arXiv:1703.05334* (2017).

11. Levine, Y., Yakira, D., Cohen, N. & Shashua, A. Deep learning and quantum physics: A fundamental bridge. *arXiv preprint arXiv:1704.01552* (2017).

12. Magesan, E., Gambetta, J. M., Córcoles, A. D. & Chow, J. M. Machine learning for discriminating quantum measurement trajectories and improving readout. *Phys. Rev. Lett.* **114**, 200501 (2015).

13. Hentschel, A. & Sanders, B. C. Machine learning for precise quantum measurement. *Phys. Rev. Lett.* **104**, 2–5 (2010).

14. Mills, K., Spanner, M. & Tamblyn, I. Deep learning and the schrödinger equation. *Phys. Rev. A* **96**, 042113 (2017).

15. Bukov, M. et al. Machine learning meets quantum state preparation. the phase diagram of quantum control. *arXiv preprint arXiv:1705.00565* (2017).

16. Chapman, R. J., Ferrie, C. & Peruzzo, A. Experimental Demonstration of Self-Guided Quantum Tomography. *Phys. Rev. Lett.* **117**, 1–5 (2016).

17. Krenn, M., Malik, M., Fickler, R., Lapkiewicz, R. & Zeilinger, A. Automated search for new quantum experiments. *Phys. Rev. Lett.* **116**, 090405 (2016).

18. Chiappetta, P., Colangelo, P., De Felice, P., Nardulli, G. & Pasquariello, G. Higgs search by neural networks at LHC. *Phys. Lett. B* **322**, 219–223 (1994).

19. Rupp, M., Tkatchenko, A., Müller, K.-R. & Von Lilienfeld, O. A. Fast and accurate modeling of molecular atomization energies with machine learning. *Phys. Rev. Lett.* **108**, 058301 (2012).

20. Rupp, M. Machine learning for quantum mechanics in a nutshell. *Int. J. Quantum Chem.* **115**, 1058–1073 (2015).

21. Biswas, R. et al. Application of machine learning algorithms to the study of noise artifacts in gravitational-wave data. *Phys. Rev. D.* **88**, 062003 (2013).

22. Ackley, D. H., Hinton, G. E. & Sejnowski, T. J. A learning algorithm for boltzmann machines. *Cogn. Sci.* **9**, 147–169 (1985).

23. Guṭă, M. & Kotłowski, W. Quantum learning: asymptotically optimal classification of qubit states. *New J. Phys.* **12**, 123032 (2010).

24. Haykin, S. S., Haykin, S. S., Haykin, S. S. & Haykin, S. S. *Neural networks and learning machines* 3 (Pearson Upper Saddle River, NJ, 2009).

25. Bell, J. S. Speakable and Unspeakable in Quantum Mechanics. *Am. J. Phys.* **57**, 567 (1989).

26. Clauser, J. F., Horne, M. A., Shimony, A. & Holt, R. A. Proposed experiment to test local hidden-variable theories. *Phys. Rev. Lett.* **23**, 880 (1969).

27. Nielsen, M. A. & Chuang, I. Quantum computation and quantum information (2002).

28. Gühne, O. & Tóth, G. Entanglement detection. *Phys. Rep.* **474**, 1–75 (2009).

29. Gurvits, L. Classical deterministic complexity of Edmonds' Problem and quantum entanglement. *Proceedings of the thirty-fifth ACM symposium on Theory of computing - STOC '03* 10 (2003).

30. Johnston, N. QETLAB: A MATLAB toolbox for quantum entanglement, version 0.9. http://qetlab.com (2016).

31. Horodecki, M., Horodecki, P. & Horodecki, R. Separability of mixed states: necessary and sufficient conditions. *Phys. Lett. A* **223**, 1–8 (1996).

32. Glorot, X., Bordes, A. & Bengio, Y. Deep Sparse Rectifier Neural Networks. *Aistats* **15**, 315–323 (2011).

33. Lu, D. et al. Tomography is necessary for universal entanglement detection with single-copy observables. *Phys. Rev. Lett.* **116**, 230501 (2016).

34. Bennett, C. H. et al. Unextendible product bases and bound entanglement. *Phys. Rev. Lett.* **82**, 5385 (1999).

35. Li, D., Li, X., Huang, H. & Li, X. Simple criteria for the slocc classification. *Phys. Lett. A* **359**, 428–437 (2006).

36. Mermin, N. D. Extreme quantum entanglement in a superposition of macroscopically distinct states. *Phys. Rev. Lett.* **65**, 1838 (1990).

37. Svetlichny, G. Distinguishing three-body from two-body nonseparability by a bell-type inequality. *Phys. Rev. D.* **35**, 3066 (1987).

38. Banaszek, K., Cramer, M. & Gross, D. Focus on quantum tomography. *New J. Phys.* **15**, 125020 (2013).

39. Freedman, S. J. & Clauser, J. F. Experimental Test of Local Hidden-Variables Theories. *Phys. Rev. Lett.* **28**, 938 (1972).

40. Giustina, M. et al. Bell violation using entangled photons without the fair-sampling assumption. *Nature* **497**, 227–230 (2013).

41. Hensen, B. et al. Loophole-free Bell inequality violation using electron spins separated by 1.3 kilometres. *Nature* **526**, 682–686 (2015).

42. Dunne, R. A. & Campbell, N. A. On the pairing of the softmax activation and cross-entropy penalty functions and the derivation of the softmax activation function. In *Proc. 8th Aust. Conf. on the Neural Networks, Melbourne,* 181, vol. 185 (1997).

43. Chollet, F. et al. Keras. https://github.com/keras-team/keras (2015).

44. Lu, S. et al. A separability-entanglement classifier via machine learning. *arXiv preprint arXiv:1705.01523* (2017).

45. Ma, Y. C. Transforming bell's inequalities into state classifiers with machine learning. https://doi.org/10.6084/m9.figshare.6231662 (2018).