Quantum Information Processing (2023) 22:341
https://doi.org/10.1007/s11128-023-04068-0

Quantifying unknown entanglement by neural networks

Xiaodie Lin[1] · **Zhenyu Chen**[1] · **Zhaohui Wei**[2,3]

Received: 10 March 2023 / Accepted: 28 July 2023

Abstract
Quantum entanglement plays a crucial role in quantum information processing tasks and quantum mechanics; hence, quantifying unknown entanglement is a fundamental task. However, this is also challenging, as entanglement cannot be measured by any observables directly. In this paper, we train neural networks to quantify unknown entanglement, where the input features for neural networks are the outcome statistics data produced by measuring target quantum states with local or even single-qubit Pauli observables, and the training labels are well-chosen quantities. For bipartite quantum states, this quantity is coherent information, which is a lower bound for many popular entanglement measures, like the entanglement of distillation. For multipartite quantum states, we choose this quantity as the geometric measure of entanglement. It turns out that the neural networks we train have very good performance in quantifying unknown quantum entanglement and can beat previous approaches like semi-device-independent protocols for this problem easily in both precision and application range. We also observe an interesting phenomenon that on quantum states with stronger quantum nonlocality, the neural networks tend to have better performance, though we do not provide them any knowledge on quantum nonlocality.

1 Introduction

It is well-known that quantum entanglement plays a central role in designing quantum information processing tasks and understanding quantum mechanics. As a result, characterizing unknown quantum entanglement is a fundamental problem, where density matrices are not available. However, as entanglement cannot be measured directly using any physical observables, this is a very challenging task that has attracted a lot of attentions.

✉ Zhaohui Wei
weizhaohui@gmail.com

1 Institute for Interdisciplinary Information Sciences, Tsinghua University, Beijing 100084, China

2 Yau Mathematical Sciences Center, Tsinghua University, Beijing 100084, China

3 Yanqi Lake Beijing Institute of Mathematical Sciences and Applications, Beijing 101407, China

Published online: 08 September 2023

Two most popular approaches adopted by quantum experimentalists for this problem are quantum tomography and entanglement witness [1–8], though both of them suffer from obvious drawbacks. Indeed, quantum tomography extracts the full information of target quantum systems, which allows us to look into the underlying entanglement. However, to achieve this exponential amount of resources is needed; thus, tomography is not a realistic option even in the realm of noisy, intermediate scale quantum (NISQ) era. On the other hand, though the idea of entanglement witness is much more efficient, it needs much nontrivial prior knowledge on target systems and usually can only provide an unreliable certification of the existence of entanglement due to imperfect quantum measurements [6]. In order to overcome the difficulties of these two approaches, device-independent protocols have been proposed to certify the existence of entanglement [9–15], where all conclusions are drawn based only on observed quantum nonlocality regardless of other implementation details and thus are reliable.

Meanwhile, as quantum computing becomes more and more important from the viewpoint of engineering, how to *quantify* unknown entanglement, an even harder task than only certifying the existence, has also been an important and urgent issue. However, even if full information of target quantum states is available, the quantification of entanglement is still a very challenging problem, especially when it comes to mixed quantum states or multipartite quantum states [16, 17]. In fact, though quite a lot of reasonable measures have been proposed, most of them are very hard to calculate. Naturally, when the density matrices of target quantum states are unknown, quantifying their entanglement is even more complicated, as one has to characterize target quantum states properly by quantum measurements first, which is usually very costly.

Among the measures of entanglement, the quantum Rényi entropies and logarithmic negativity are relatively easier to handle [16, 18, 19] and thus have been extensively studied in recent years. Indeed, some interesting protocols have been proposed to characterize these two measures by analyzing the data produced by well-designed measurements [20–23]. Particularly, machine learning tools like neural networks have been introduced to quantity logarithmic negativity, which provides very nice performance [24] (see also Ref. [25]). However, it should be pointed out that, in addition to the quantum Rényi entropies and logarithmic negativity, we have many more important measures of entanglement, like the entanglement of distillation [16]. Undoubtedly, proposing protocols that are able to experimentally quantify these widely used measures of entanglement is a very important task.

In fact, device-independent protocols have also been utilized to quantify these entanglement measures for unknown quantum states [26–31]. Particularly, by defining a concept called *nondegenerate Bell inequalities*, a semi-device-independent (semi-DI) scheme that can lower bound many important entanglement measures was found [32], including the entanglement of distillation mentioned above, which is usually a lower bound for other entanglement measures [33]. Later, this approach has been generalized to quantify multipartite entanglement by the geometric measure and the relative entropy of entanglement [34].

However, the results provided by device-independent protocols like those in Refs. [32, 34] are very conservative and weak. In fact, usually these protocols can provide informative results only when the observed nonlocality is close to the strongest, which

seriously restricts their applications. As a result, new schemes that can experimentally quantify these entanglement measures with better performance and wider application ranges have to be explored, which is also the main motivation of the current paper.

In this paper, we apply neural networks to quantify unknown quantum entanglement by several popular entanglement measures. Specifically, to train the neural networks, we choose the outcome statistics data produced by locally measuring sample quantum states as features, which is relatively easy to obtain experimentally. As a comparison, the neural networks in Ref. [24] choose partial transposed moments as the input feature, which usually need exponential numbers of measurements to compute [23, 35].

Meanwhile, we choose coherent information for bipartite quantum states, or the geometric measure of entanglement for multipartite quantum states, as feature labels or prediction targets. Note that for bipartite quantum states, though coherent information is actually not an entanglement measure, it is a lower bound for many important entanglement measures, like the entanglement of distillation [36]; thus, our results can be regarded as an indirect approach to quantify these entanglement measures. By running the trained neural networks on new unknown quantum states, we show that they have very good performance in quantifying the involved entanglement. For example, the prediction results can beat those of known device-independent protocols for this problem easily in both precision and application range, implying the high potential value of neural networks in quantifying unknown quantum entanglement. Furthermore, we observe interesting phenomenon that neural networks have better performance on quantum states with stronger nonlocality, though our training data do not contain any direct information on quantum nonlocality.

2 Settings and entanglement measures

As mentioned above, the data features that neural networks work with are the measurement outcome statistics produced by local measurements on target quantum states. Suppose we have a physical system shared by n space-separated parties, and the state is ρ. Each party i, where $i \in \{1, \ldots, n\}$, has a set of measurement devices labeled by X_i and the possible measurement outcomes are labeled by A_i. Assume every party randomly chooses a measurement $x_i \in X_i$ to measure their subsystems and record the corresponding outcomes $a_i \in A_i$. After repeating the procedure for many times, they have the joint probability distribution $p(a_1 a_2 \ldots a_n | x_1 x_2 \ldots x_n)$, which indicates the probability of obtaining outcomes $a_i \in A_i$ when measurements $x_i \in X_i$ are chosen. Suppose $M_{x_i}^{a_i}$ is the measurement operator with outcome a_i for the measurement device x_i performed by the party i, then it holds that

$$p(a_1 a_2 \ldots a_n | x_1 x_2 \ldots x_n) = \mathrm{Tr}\left(\left(\bigotimes_{i=1}^{n} M_{x_i}^{a_i}\right)\rho\right). \tag{1}$$

Note that although our setting is quite similar with Bell experiments, we will not talk about nonlocality and Bell inequalities at most time, though like in Bell experiments

we hope that neural networks will reveal us nontrivial information on the amount of entanglement based only on the data $p(a_1a_2\ldots a_n|x_1x_2\ldots x_n)$.

For this, we have to choose proper entanglement measures for our tasks. Before doing that, let us first recall briefly how neural networks, a machine learning model we will utilize in this paper, work. First, we set up a proper mathematical model for the target problem, which usually contains one input layer, zero or more hidden layers, and one output layer, and every two adjacent layers involve a lot of parameters. Then we gather the features of many representative sample data with correct labels, called training data, and input the features into the model. We hope the outputs of the model are as consistent as possible with the correct labels, for which we constantly adjust the parameters by proper optimization methods, and this process is called training. If the training dataset is representative and the model is well trained, eventually the well-tuned model will capture crucial properties of the training dataset and can predict the labels of a new test dataset unseen before very precisely.

Accordingly, in our task the inputs of neural networks, or the features of training and test datasets composed of the quantum states, are chosen to be the outcome statistics data generated via Eq. (1), and the outputs of neural networks, or the labels of training and test datasets, are chosen to be target entanglement measures we are interested in. As a consequence, in order to prepare training sets with valid labels, we have to make sure that the corresponding entanglement measures can be computed with high precision on all training quantum states. As is well-known, this is a very challenging task. For example, if we choose the entanglement of distillation as our target entanglement measure, this requirement is very hard to satisfy. Therefore, here we have to make proper compromises.

For bipartite quantum states, a possible way to compromise is to quantify coherent information instead. Coherent information is a fundamental quantity responsible for the capability of transition of quantum information [37, 38]. For a quantum state ρ acting on a composite Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B$, its coherent information is defined as

$$I_C(\rho) = S(\rho_A) - S(\rho), \tag{2}$$

where $\rho_A = \mathrm{Tr}_B(\rho)$ is the subsystem in $\mathcal{H}_A$ and $S(\rho)$ is the Von Neumann entropy of ρ. In fact, the coherent information is not an entanglement measure. However, it is relatively easy to calculate, and most importantly, it turns out that [36]

$$E_D(\rho) \geq I_C(\rho), \tag{3}$$

where $E_D(\rho)$ is the entanglement of distillation. Therefore, if we choose the prediction target of neural networks as coherent information, the labels of training data should be easy to gather, and if the model has good performance, we can eventually obtain very nontrivial information on $E_D(\rho)$ (and many other entanglement measures) for target quantum states.

We now turn to the case of multipartite quantum states. Unfortunately, in this case very few well-known entanglement measures or quantities related to entanglement measures can be calculated accurately, which makes it very hard to prepare representative training data with valid labels for our neural networks. In this paper, we choose

the geometric measure of entanglement as our target entanglement measure. The reason is that by numerical methods, we can barely manage to calculate the entanglement measure of some pure multipartite quantum states, which serve as the training data.

Recall that the geometric measure of entanglement is firstly introduced in the setting of bipartite pure states [39] and then generalized to the multipartite setting [40]. For a n-partite pure state $|\psi\rangle$, the geometric measure of entanglement is defined as

$$E_G(|\psi\rangle) = 1 - G(|\psi\rangle)^2, \tag{4}$$

where

$$G(|\psi\rangle) = \sup_{|\phi\rangle \in \text{sep}_n} |\langle\psi|\phi\rangle|, \tag{5}$$

and sep_n is the set of n-partite product pure states. For a mixed state ρ, the geometric measure of entanglement is defined by convex roof construction, which is

$$E_G(\rho) = \min_{\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|} \sum_i p_i E_G(|\psi_i\rangle). \tag{6}$$

3 Quantifying bipartite entanglement

3.1 Qutrit-qutrit case

As a warm-up, we first consider quantifying the coherent information of a special class of qutrit-qutrit quantum states, which have the form

$$\rho_{3_\epsilon}^{mv} = (1-\epsilon) I_9/9 + \epsilon \left|\psi_3^{mv}\right\rangle\left\langle\psi_3^{mv}\right|, \tag{7}$$

where $\epsilon \in [0, 1]$ and $|\psi_3^{mv}\rangle = \gamma(|00\rangle + |22\rangle) + \sqrt{1-2\gamma^2}|11\rangle$ with $\gamma \approx 0.617$. Note that $|\psi_3^{mv}\rangle$ is the quantum state that violates the Collins-Gisin-Linden-Massar-Popescu (CGLMP) inequality maximally [41–44]. The reason why we choose this class of states as our first target is that their coherent information can also been quantified using the semi-DI approach proposed in Ref. [32], which provides us a nice chance to compare the performance of these two different approaches.

As mentioned above, the sample features for the neural networks we choose are the outcome statistics data produced by locally measuring the involved quantum states, which has a form of Eq. (1). Here the local measurements we choose are the ones that maximize the violation of the CGLMP inequality, and it turns out that this choice can achieve very good performance. Furthermore, later we will see that other choices of local measurements may also provide similar performance (see also Ref. [45]). Specifically, these local measurements we choose can be described by observables A_a

and B_b with the eigenvectors

$$|i\rangle_{A,a} = \frac{1}{\sqrt{d}} \sum_{k=0}^{d-1} \exp\left(\mathbf{i}\frac{2\pi}{d}k(i+\alpha_a)\right)|k\rangle_A \tag{8}$$

and

$$|j\rangle_{B,b} = \frac{1}{\sqrt{d}} \sum_{l=0}^{d-1} \exp\left(\mathbf{i}\frac{2\pi}{d}l(-j+\beta_b)\right)|l\rangle_B \tag{9}$$

respectively [41, 46], where $a, b \in \{1, 2\}$, $i, j \in \{0, 1, \ldots, d-1\}$, $\mathbf{i} = \sqrt{-1}$ is the imaginary number, and the phases read $\alpha_1 = 0, \alpha_2 = 1/2, \beta_1 = 1/4$ and $\beta_2 = -1/4$. In this subsection, we let $d = 3$.

To generate the training dataset, a number of bipartite quantum states of form $(1-\epsilon)I_9/9 + \epsilon\rho_0$ are sampled. Here we select ϵ from 0.2 to 1 at intervals 0.025 and ρ_0 as arbitrary quantum state in $\mathcal{H}^3 \otimes \mathcal{H}^3$. Specifically, we randomly generate ρ_0 of the form

$$\rho_0 = \sum_{i=0}^{k-1} \lambda_i |u_i\rangle\langle u_i|, \tag{10}$$

where $k \in \{1, 2, \ldots, 9\}$ is uniformly picked. For this, we first randomly choose positive λ_i's such that they satisfy $\sum_i \lambda_i = 1$. Then we produce a Haar random unitary matrix, and let $|u_i\rangle$ be its i-th column (therefore $\{|u_i\rangle\}$ is a set of orthonormal basis). Through this approach, we can randomly sample a large amount of quantum states from the whole quantum state space, which allows us to prepare representative training datasets.

For each ϵ not larger than 0.75, 2000 states are randomly sampled, while for the other values of ϵ, we increase the number of samples to 4000. Then, the corresponding outcome statistics data, which has a form of Eq. (1) and a dimension 36 (each party has 2 measurement settings with 3 outcomes), are recorded when measuring A_a and B_b locally on the sampled states. Together with the corresponding coherent information as the labels, 86,000 instances as training dataset with dimension 36 are fed to a 4-hidden-layer fully connected neural network with 400, 200, 100, 50 neurons in each layer, respectively.

The mean squared error (MSE) is a widely used metric to quantify the performance of the model, which is defined as

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}\left(Y_i - \hat{Y}_i\right)^2, \tag{11}$$

where Y_i is the exact value of coherent information, $\hat{Y}_i$ is the predicted value, and N is the size of the test dataset. In our models, we use the MSE as the loss function for training.

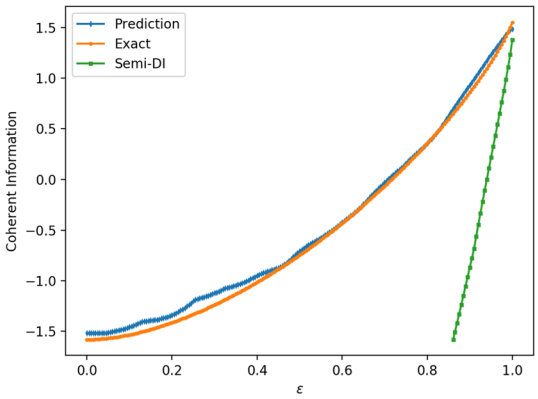

Fig. 1 The neural network predictions for the coherent information of $\rho_{3_\epsilon}^{mv}$. The blue line and the orange line represent the prediction values and the exact values, respectively, where the MSE of these prediction values is 0.0040. In addition, the green line shows the lower bound for the coherent information provided by the semi-DI approach in Ref. [32], which has a much narrower application range (Color figure online)

After the training stage ends, we apply our model to predict the coherent information of a test dataset of form $\rho_{3_\epsilon}^{mv}$, which is generated by selecting ϵ from 0 to 1 at intervals 0.005. Here we would like to point out that $\rho_{3_\epsilon}^{mv}$ does not appear in our training dataset. The prediction results are shown in Fig. 1, in which the blue line and the orange line represent the prediction and the exact values of coherent information, respectively. It can be seen that our model behaves very well in this task, where the residuals are quite small, and the MSE is 0.0040.

As a comparison, the above setting can also be handled by the semi-DI approach given by Ref. [32]. As illustrated by the green line in Fig. 1, the method in Ref. [32] can provide nontrivial lower bound for the coherent information only when ϵ ranges from 0.8610 to 1, and the residuals are also much larger. Therefore, in this case the neural network approach beats the approach in Ref. [32] easily.

We now apply neural networks to quantify the coherent information for more general qutrit-qutrit quantum states. For this, we consider the following four different classes of quantum states.

- $(1-\epsilon)I_9/9 + \epsilon|\psi\rangle\langle\psi|$, where $|\psi\rangle$ is an arbitrary pure state (called *pure states under white noise*).
- $(1-\epsilon)I_9/9 + \epsilon\rho_0$, where ρ_0 is an arbitrary mixed state (called *mixed states under white noise*).
- $|\psi\rangle\langle\psi|$, where $|\psi\rangle$ is an arbitrary pure state (called *general pure states*).
- ρ_0, where ρ_0 is an arbitrary mixed states (called *general mixed states*).

In order to make the training data as representative as possible, we will generate them by picking up sample states from all the four classes. After the training stage is finished, we then test the performance of our neural network for each class separately. The neural network we choose has the same structure as before, whose 4 hidden layers has 400, 200, 100, 50 neurons, respectively.

More concretely, we first use the same way as in Eq. (10) to sample ρ_0 or $|\psi\rangle$ (by setting $k = 1$ there, we get a pure state $|\psi\rangle$). Next, when generating training data from the classes of pure states under white noise and mixed states under white noise, we select ϵ at intervals 0.025 and randomly generate 2000 states for each ϵ from 0.2 to

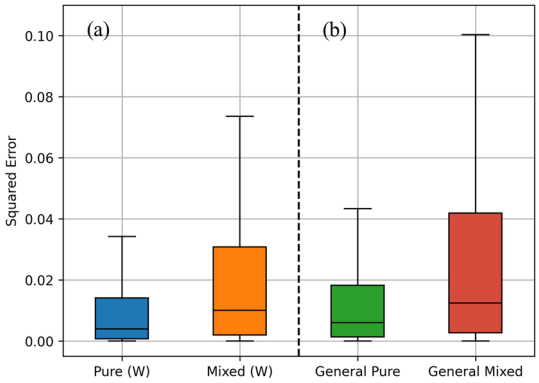

Fig. 2 The neural network predictions for the coherent information of qutrit-qutrit quantum states. Each boxplot indicates the degree of dispersion and skewness in the squared errors for each case. **a** Boxplots of pure states under white noise (blue, or 'Pure (W)') and mixed states under white noise (orange, or 'Mixed (W)'), where the corresponding MSEs are 0.0137 and 0.0281, respectively. **b** Boxplots of general pure states (green) and general mixed states (red), where the corresponding MSEs are 0.0161 and 0.0407, respectively (Color figure online)

0.75, and 4000 states for each ϵ from 0.775 to 0.975, where 82,000 sample states are picked up for each class.

When generating training data from the classes of general pure and general mixed states, whose coherent information considered for training dataset ranges from 0 to $\log_2 3$ and from $-\log_2 3$ to 1.5, respectively, we choose to keep or drop the sampled ρ_0 or $|\psi\rangle$ according to their coherent information, aiming to select 4000 states randomly in each interval for coherent information of size 0.1. Here the only exception is for general mixed states with coherent information larger than 1.5. Due to the low sampling efficiency, we do not insist on the number of states produced in this case and directly collect all such states during the sampling process. Eventually, 64,000 general pure states, 124,211 general mixed states, and 164,000 pure states under white noise and mixed states under white noise are chosen as training states. Performing the same local measurements as the previous experiment on them and recording the outcome statistics data for each state as the features, together with the corresponding coherent information as the label, we obtain the whole training dataset, which has 352,211 sample states.

After training the model, we now use the same model to test its performance on the above four classes of quantum states separately. That is to say, for each class we generate a test dataset by sampling around 2000 random states uniformly according to coherent information. Similarly, we use the MSE to measure the prediction performance of our model. The results are listed in Fig. 2. For convenience, we represent the experimental results in the form of boxplot, where each boxplot displays a dataset based on a five-number summary: the lowest data point excluding any outliers (minimum), the median of the lower half of the dataset (first quartile), the median of the dataset, the median of the upper half of the dataset (third quartile), and the largest data point excluding any outliers (maximum).

As we can see in Fig. 2a, when predicting the coherent information of pure states under white noise and mixed states under white noise, the MSEs are 0.0137 and 0.0281, respectively. Here the distribution of the squared error for pure states under white noise is very narrow, indicating the very high-quality predictions. Figure 2b shows the results for general pure states and general mixed states, where the corresponding MSEs are 0.0161 and 0.0407, respectively. Overall, it can be said that even for the most general case of qutrit-qutrit quantum states, the neural network model we trained still enjoys decent performance in predicting coherent information.

3.2 Higher-dimensional cases

We now utilize neural networks to predict the coherent information of bipartite quantum states in higher-dimensional Hilbert space $\mathcal{H}^d \otimes \mathcal{H}^d$. In this case, the task of picking up proper training dataset is more challenging. In order to make the sample states for training as representative as possible, we consider a class of quantum states of the form

$$\rho_d = \alpha\rho_0 + \beta I_{d^2}/d^2 + \gamma|\psi_d^{me}\rangle\langle\psi_d^{me}|, \tag{12}$$

where $\alpha, \beta, \gamma \geq 0$, $\alpha + \beta + \gamma = 1$, $\rho_0 \in \mathcal{H}^d \otimes \mathcal{H}^d$ is a random quantum state generated according to Eq. (10), I_{d^2}/d^2 is the bipartite maximally mixed state, and $|\psi_d^{me}\rangle = \sum_{i=0}^{d-1} |ii\rangle/\sqrt{d}$ is the maximally entangled state. In other words, ρ_d is a linear combination of three completely different types of quantum states.

To demonstrate how to generate training dataset for this case, let us take $d = 5$ as an example. The training states will be composed of two parts. Firstly, we generate the first part of the training states by sampling ρ_5 in Eq. (12), where about 170,000 states are chosen. More concretely, when α is no less than 0.4, α and β are varied at intervals 0.02, and for each choices of the coefficients, 200 random ρ_5 are produced by sampling ρ_0. When α is smaller than 0.4, we vary α, β at intervals 0.04 and sample 100 ρ_5 states similarly for each choices of the coefficients. The second part of the training states will be produced in a similar way with the qutrit-qutrit case. That is, we separately sample 82,000 pure states under white noise, 130,000 mixed states under white noise, 40,000 general pure states, and 84,000 general mixed states in such a way that their coherent information is distributed evenly. Here, when sampling general pure states, we find that it is hard to sample states with coherent information less than 0.8. In order to have a relatively complete coherent information distribution, we generate general pure states by

$$|\psi_5\rangle \propto \alpha|\psi_5'\rangle + (1-\alpha)|\psi_5^{sep}\rangle, \tag{13}$$

where $\alpha \in [0, 1]$, $|\psi_5'\rangle$ and $|\psi_5^{sep}\rangle$ are a random pure state and a random separable pure state in $\mathcal{H}^5 \otimes \mathcal{H}^5$, respectively. Again, random pure states are generated by setting $k = 1$ in Eq. (10) and each random separable pure state is a tensor product of two random pure states in $\mathcal{H}^5$. Our previous experiences show that the local measurements given in Eqs. (8) and (9) can produce informative measurement outcome statistics

Table 1 The structure and configuration details of the convolutional neural network (see Refs. [47, 48] for introductions to convolutional neural networks)

Layers	Type	Neurons	Filters	Kernel size	Strides	Pool size
0–1	Convolution2D	(None, 9, 9, 32)	32	2×2	1 × 1	-
1–2	Max-pooling2D	(None, 8, 8, 32)	–	–	1 × 1	2×2
2–3	Convolution2D	(None, 7, 7, 64)	64	2×2	1 × 1	-
3–4	Max-pooling2D	(None, 6, 6, 64)	–	–	1 × 1	2×2
4–5	Convolution2D	(None, 5, 5, 64)	64	2×2	1 × 1	-
5–6	Fully connected	(None, 32)	–	–	–	–
6–7	Fully connected	(None, 1)	–	–	–	–

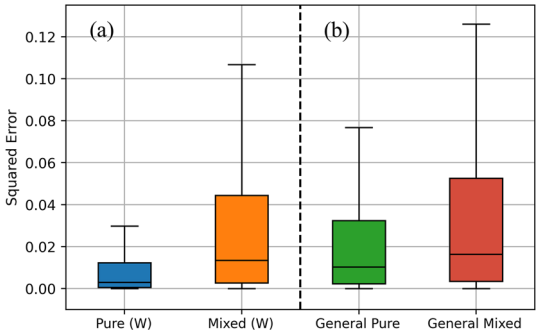

Fig. 3 The neural network predictions for the coherent information of higher-dimensional quantum states, where $d = 5$. Each boxplot indicates the degree of dispersion and skewness in the squared errors for each case. **a** Boxplots of pure states under white noise (blue, or 'Pure (W)') and mixed states under white noise (orange, or 'Mixed (W)'), where the corresponding MSEs are 0.0123 and 0.0356, respectively. **b** Boxplots of general pure states (green) and general mixed states (red), where the corresponding MSEs are 0.0275 and 0.0436, respectively (Color figure online)

data. Consequently, after picking up the states for training, we measure them by these local measurements again. The resulting measurement outcome statistics data, whose dimension is now 100, serves as the features for the neural network. As training dataset, we compute the coherent information for each sample state as the corresponding label.

Compared with fully connected neural networks, in higher-dimensional cases we find that convolutional neural networks have better performance in quantifying entanglement (see Refs. [47, 48] for introductions to convolutional neural networks). The reason for this behavior may come from the increasing dimension of the input data, since convolutional neural networks are good at extracting information from high-dimensional data. The structure of the convolutional neural network we choose is listed in Table 1 . Again, we test the performance of the trained neural network by utilizing it to separately predict the coherent information of four classes of quantum states, that is, pure states under white noise, mixed states under white noise, general pure states and general mixed states.

For each class of test states, we fix the number of instances to be 2000, and the experimental results are presented in Fig. 3. It can be seen that the overall performance

Table 2 The MSEs of predicting coherent information for different dimensions

Dimension	$d = 5$	$d = 6$	$d = 7$	$d = 8$	$d = 9$	$d = 10$
Pure	0.0135	0.0110	0.0107	0.0090	0.0093	0.0086
Mixed	0.0347	0.0321	0.0255	0.0217	0.0171	0.0149

Each case is trained with around 40,000 samples and tested with over 2000 samples

is similar to the qutrit-qutrit case, and the MSEs are 0.0123, 0.0356, 0.0275, and 0.0436, respectively.

In addition to the boxplots, to further describe the degree of dispersion of the prediction results, we also compute the relative error of predictions for test states. For the class of general mixed states, which is the hardest to predict, the mean relative error is smaller than 15% (10%) for test states whose absolute values of the exact coherent information are larger than 0.5 (1.0). We stress that all these results are predicted by one single trained model. Hence, it is fair to say that our model has decent performance even for general quantum states.

To examine the potential power of neural networks better, we now increase the dimension of target quantum states even further and use a convolutional neural network similar to the one in Table 1 to predict coherent information, where the only difference is that we increase the kernel size and pool size from 2×2 to 3×3. Note that when quantum dimension goes up, it is very challenging to handle the sharply increasing size of the training set. Therefore, from now on we only do the following two experiments: (1) Both the training states and the test states are sampled from general mixed states; (2) both the training states and the test states are sampled from general pure states. In each case, the number of training samples is around 40,000, and the number of test samples is over 2000. Similar to before, we try to pick up these samples evenly according to their coherent information.

We run the model for quantum states with dimensions from 5 to 10, and the results are listed in Table 2. Compared with Fig. 3b, the prediction MSEs for general pure states and general mixed states are improved from 0.0275 to 0.0135 and from 0.0436 to 0.0347, respectively, showing that utilizing prior knowledge on the type of target quantum states can improve the performance. Furthermore, a very surprising fact is that, unlike fully connected neural networks, when the quantum dimension goes up, the convolutional neural network model behaves even better, providing more evidence showing that the convolutional neural network is a more suitable choice for higher-dimensional cases.

In addition, to further evaluate the performance of our models, we conduct uncertainty estimates to characterize the distribution of residuals in the prediction results of our models. For this, we utilize a technique called bootstrapping [49, 50]. Specifically, we split a large dataset composed of quantum states into k parts of the same size, where $k - 1$ parts combined are used for training and the remaining one for testing. In this way, we obtain k different models for the same task and thus have a chance to determine whether our approach has stable performance. Here we let $k = 100$. For each model of the k ones, after training we use it to predict coherent information for the corresponding test dataset and then record the corresponding residuals. Putting all

the residual data from the k models together, we calculate the 5% quantile and the 95% quantile, which allows us to get the 90% confidence interval for the predictions of these models. The results are presented in Table 3.

In Table 3, we can see that if the model predicts the coherent information as v for an unknown 5-dimensional pure state, the accurate coherent information lies within the interval $[v - 0.1872, v + 0.1815]$ with probability 90%. Overall, the behavior of the confidence intervals for different dimensions has a similar pattern to that of the MSE. For example, the confidence intervals of general pure states are narrower than that of general mixed states. Remarkably, it turns out that the confidence intervals become narrower as the dimension increases, implying that the performance of our model will not decay when the size of target quantum states becomes larger.

Lastly, to exploit the flexibility of our approach, we now replace the measurements given in Eqs. (8) and (9) with tensor products of single-qubit Pauli observables. This means that in order to produce outcome statistics data as in Eq. (1), measuring involved quantum states becomes much easier. Specifically, Alice now measures all quantum states by the observables

$$A_1' = \sigma_Z^{\otimes k} \quad A_2' = \sigma_X^{\otimes k} \tag{14}$$

and Bob measures by

$$B_1' = \frac{(-\sigma_Z - \sigma_X)^{\otimes k}}{2\sqrt{2}} \quad B_2' = \frac{(\sigma_Z - \sigma_X)^{\otimes k}}{2\sqrt{2}}, \tag{15}$$

where σ_X, σ_Z are Pauli matrices and $k = \lceil \log_2 d \rceil$. When $2^k > d$, quantum states in $\mathcal{H}^d \otimes \mathcal{H}^d$ can be embedded to a $\mathcal{H}^{2^k} \otimes \mathcal{H}^{2^k}$ Hilbert space by padding zeros to the redundant bases. In this way, each observable has 2^k outcomes. Here, we repeat the previous experiments, where all experimental settings keep the same except the measured observables. The corresponding prediction results are presented in Tables 4 and 5.

By introducing new observables, the number of measurement outcomes becomes 2^k rather than d. This makes the outcome statistics data more informative when $2^k > d$, and as a result, for general mixed states the prediction quality is improved (the MSEs are lower and the confidence intervals are narrower). For the special case of $d = 8$, where $2^k = d$, for general mixed states the two different approaches have the same MSE, which is 0.0217, and two similar confidence intervals. As a comparison, for general pure states the performance of Pauli observables is not as good as the measurements given in Eqs. (8) and (9), but still comparable. Furthermore, despite the different behaviors for general pure states and general mixed states, both of the two approaches show the same trend that the performance becomes even better as the dimension increases. Therefore, the results in Tables 4 and 5 imply clearly that the performance of our neural networks is still very stable when different measurement settings are chosen.

Table 3 90% confidence intervals of the residuals of predicting coherent information for different dimensions

Dimension	$d = 5$	$d = 6$	$d = 7$	$d = 8$	$d = 9$	$d = 10$
Pure	[−0.1872, 0.1815]	[−0.1679, 0.1645]	[−0.1578, 0.1615]	[−0.1607, 0.1720]	[−0.1552, 0.1591]	[−0.1553, 0.1663]
Mixed	[−0.2986, 0.3275]	[−0.2863, 0.3068]	[−0.2588, 0.2560]	[−0.2467, 0.2315]	[−0.2219, 0.2286]	[−0.2042, 0.2195]

Table 4 The MSEs of predicting coherent information with outcome statistics data obtained by qubit Pauli observables

Dimension	$d = 5$	$d = 6$	$d = 7$	$d = 8$	$d = 9$	$d = 10$
Pure, local	0.0165	0.0195	0.0243	0.0216	0.0173	0.0144
Mixed, local	0.0278	0.0255	0.0221	0.0217	0.0141	0.0110

Combining all the evidence above together, we can see that the neural networks we trained enjoy a very decent performance in quantifying unknown bipartite entanglement, even if the quantum dimension is high.

4 Quantifying multipartite entanglement

Compared with bipartite entanglement, multipartite entanglement has much more complicated mathematical structures. Many different measures have been defined to quantify multipartite entanglement, and almost all of them are hard to calculate. As mentioned before, in order to apply neural network models to quantify unknown multipartite entanglement, we have to make sure the labels, which is some entanglement measure here, of training states are computationally tractable. For this, we choose the geometric measure of entanglement as our target entanglement measure, as it is relatively easy to handle among well-known multipartite entanglement measures, especially for the case of pure states. Therefore, in this section we will only train and test our neural networks using three-qubit pure states.

To prepare training data, again every random three-qubit pure state $|\psi\rangle \in \mathcal{H}^2 \otimes \mathcal{H}^2 \otimes \mathcal{H}^2$ is generated by Eq. (10) with $k = 1$. Then, we measure each qubit by Pauli observables σ_x and σ_y and record the outcome statistics data as the corresponding features. For the label of the above features, we first parameterize the 3-partite product pure state $|\phi\rangle$ in Eq. (5) as

$$|\phi\rangle = \bigotimes_{i=1}^{3} \left(\cos \frac{\theta_i}{2} |0\rangle + e^{\mathbf{i}\gamma_i} \sin \frac{\theta_i}{2} |1\rangle \right), \tag{16}$$

and then compute the geometric measure of entanglement $E_G(|\psi\rangle)$ by maximizing Eq. (5) over the parameters $\{\theta_i, \gamma_i \in \mathbb{R}\}$ with the L-BFGS-B method [51, 52] implemented by SCIPY [53]. The whole training dataset is generated by sampling states according to their geometric measure of entanglement. Concretely, when the value of the geometric measure of entanglement is from 0 to 0.55, we sample 5000 three-qubit pure states for each interval for the geometric measure of entanglement of size 0.1. In addition, we also sample 5000 separable states as part of the training states, and when generating measurement outcome statistics data, the involved local measurements are the same as before. Combining the above two parts, we get the training dataset with 35,000 instances. Then we use the training data to train a 4-hidden-layer fully connected neural network, where each layer has 50, 20, 10, 5 neurons, respectively.

Table 5 90% confidence intervals of the residuals of predicting coherent information with outcome statistics data obtained by qubit Pauli observables

Dimension	$d = 5$	$d = 6$	$d = 7$	$d = 8$	$d = 9$	$d = 10$
Pure, local	[−0.2156, 0.2238]	[−0.2360, 0.2422]	[−0.2599, 0.2435]	[−0.2626, 0.2559]	[−0.2178, 0.2231]	[−0.2223, 0.2213]
Mixed, local	[−0.2727, 0.2670]	[−0.2498, 0.2632]	[−0.2393, 0.2392]	[−0.2288, 0.2220]	[−0.2053, 0.2116]	[−0.2032, 0.2148]

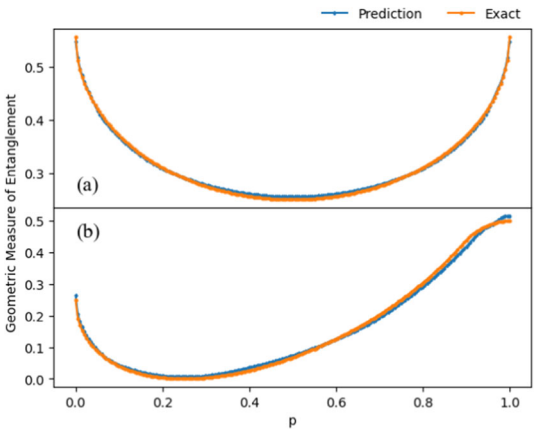

Fig. 4 The neural network predictions for the geometric measure of entanglement of three-qubit pure states. The blue line and the orange line represent the prediction values and the exact values, respectively. **a** The results for the state $|\varphi_p\rangle$, where the MSE of the prediction values is 1.26×10^{-5}. **b** The results for the state $|\varphi'_p\rangle$, where the MSE of the prediction values is 7.72×10^{-5} (Color figure online)

We now turn to the sampling of test states. Recall that there are only two classes of genuinely tripartite entanglement which are inequivalent under local operations and classical communication (LOCC) [54]. One class is represented by the Greenberger-Horne-Zeilinger (GHZ) state [55],

$$|\mathrm{GHZ}\rangle = (|000\rangle + |111\rangle)/\sqrt{2}, \tag{17}$$

and the other class is represented by the W state [54],

$$|\mathrm{W}\rangle = (|001\rangle + |010\rangle + |100\rangle)/\sqrt{3}. \tag{18}$$

Therefore, we are interested in the behavior of our model on these two states. For this, we consider the following two classes of states,

$$|\varphi_p\rangle = \sqrt{p}|\mathrm{W}\rangle + \sqrt{1-p}|\overline{\mathrm{W}}\rangle, \tag{19}$$

and

$$|\varphi'_p\rangle = \sqrt{p}|\mathrm{GHZ}\rangle + \sqrt{(1-p)/2}\left(|\mathrm{W}\rangle + |\overline{\mathrm{W}}\rangle\right), \tag{20}$$

where $p \in [0, 1]$ and $|\overline{\mathrm{W}}\rangle = (|110\rangle + |101\rangle + |011\rangle)/\sqrt{3}$.

We sample states within the classes $|\varphi_p\rangle$ and $|\varphi'_p\rangle$ to serve as our test samples, where p is chosen from 0 to 1 at intervals 0.005. The prediction results are illustrated in Fig. 4, in which (a) is for the state $|\varphi_p\rangle$ and (b) is for the state $|\varphi'_p\rangle$. As we can see, the prediction values fit the exact values very well in both cases, whose MSEs are 1.26×10^{-5} and 7.72×10^{-5}, respectively.

In addition, we also test the performance of the above-trained neural network with another test dataset composed of around 2000 random three-qubit pure states, whose geometric measure of entanglement is numerically calculated and ranges from 0 to 0.55 evenly. Again, the performance is excellent and the MSE is only 0.0011. These

high-quality predictions strongly imply that neural networks can have satisfactory performance in quantifying unknown multipartite entanglement.

5 Nonlocality facilitates neural networks

In our neural network models, the features are chosen as statistics data of the outcomes when local measurements are taken on samples. Actually this kind of statistics data is exactly what people utilize in studying nonlocality, though as stressed before, quantum nonlocality is not considered in the training and testing of our neural networks at all. However, due to this similarity, we may wonder, does quantum nonlocality play any role in the workings of our neural networks? Interestingly, we observe strong evidences, suggesting that the answer to this question is positive.

For this, let us go back to the neural network we have already trained in Sect. 3.1, which aims at quantifying coherent information for the state $\rho_{3_\epsilon}^{mv}$ defined in Eq. (7). Without further training, we run the same neural network on the following special class of quantum states,

$$\rho_{p,\gamma} = (1-p)I_9/9 + p|\phi_\gamma\rangle\langle\phi_\gamma|, \tag{21}$$

where

$$|\phi_\gamma\rangle = \gamma(|00\rangle + |11\rangle) + \sqrt{1-2\gamma^2}|22\rangle, \tag{22}$$

$\gamma \in [0.6, \sqrt{2}/2]$ and p is chosen such that the coherent information of $\rho_{p,\gamma}$ is positive. Again, for each state in the $\rho_{p,\gamma}$ class, we measure it with observables A_a and B_b given by Eqs. (8) and (9) to generate the feature and then input it into the neural network. Meanwhile, the violation of this test state to the CGLMP inequality is also computed, which as usual is used to measure quantum nonlocality. Then, after running the neural network on these test samples, we have the chance to examine whether the performance of prediction has any relation with quantum nonlocality.

The results are listed in Fig. 5. Here each blue dot represents a quantum state, which is a quantum state sampled from the $\rho_{p,\gamma}$ class. Along the x-axis, in Fig. 5a we record the exact coherent information, and in Fig. 5b we record the violation. Here the squared error is the squared gap between the exact and the predicted coherent information. It can be seen clearly that when a test state has stronger quantum nonlocality, the neural network tends to predict its coherent information more accurately.

To further analyze the statistical connections between nonlocality, coherent information, and squared error, we utilize the Pearson correlation coefficient (PCC) to analyze the data, which is a measure of linear correlation between the paired data $\{(x_1, y_1), \ldots, (x_n, y_n)\}$ defined as

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}, \tag{23}$$

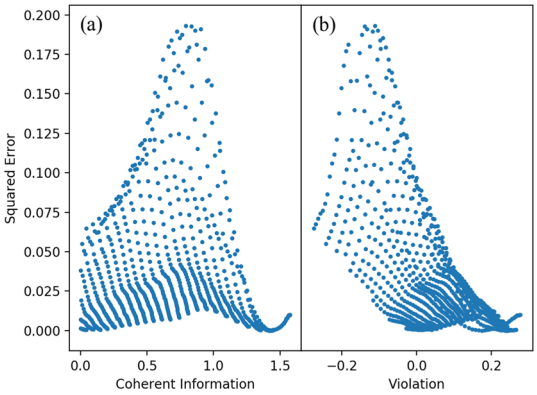

Fig. 5 The relations between coherent information, nonlocality, and squared error. Each blue dot represents a quantum state $\rho_{p,\gamma}$ with parameter pair (p, γ). **a** The relation between squared error and coherent information, where the PCC is 0.0010. **b** The relation between squared error and violation, where the PCC is −0.6041 (Color figure online)

where $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ and analogously for $\bar{y}$. We find that the PCC between squared error and coherent information is 0.0010, implying that almost no linear correlation exists between them. As a sharp difference, the PCC between squared error and violation is −0.6041, confirming that an apparent negative correlation exists between them, that is, when the violation becomes larger, the squared error shrinks.

This phenomenon could be explained as follows. Notice that the inputs of the neural network are the measurement outcome statistics generated by locally measuring two different measurement devices. In all the experiments we listed above, the statistics data are not tomographically complete to pin down a quantum state. Therefore, there usually exist different quantum states admitting the same measurement outcome statistics data, and our numerical calculations confirm this. When the measurement outcome statistics data correspond to weak nonlocality, the range of quantum states that can produce the current measurement outcome statistics is quite large. As a result, when we quantify the entanglement for a quantum state based on such measurement outcome statistics data, the output should not be a value, but an interval. Inspired by the concept of self-testing [56, 57], it can be conceived that when a measurement outcome statistics corresponds to strong nonlocality, this interval will be small, which is also indicated in Fig. 5.

Meanwhile, we stress that our motivation in this work is introducing deep-learning techniques to quantify unknown quantum entanglement. We hope to provide a realistic and economical method for future quantum industries. The outputs of our models have to be specific values of target entanglement measures, and therefore, a minor inaccuracy of these outputs is inevitable. However, we observe that a typical output of our models is roughly the middle point of the interval that the corresponding measurement outcome statistics data correspond to, which means that the results are still very informative when the interval is relatively small, as confirmed over and over by the numerical experiments we have reported. Furthermore, from the viewpoint of quantum engineering, a rough and easy-to-implement quantification for unknown entanglement is often very valuable. For example, when selecting quantum states as raw materials for a quantum factory that distills entanglement, applying the neural network approach we introduced here is a very reasonable choice, as once trained, running the neural network to quantify unknown entanglement will be very efficient.

6 Discussions

In this paper, we have showed strong evidences implying that neural networks can be used to quantify unknown quantum entanglement after being properly trained, where the performance is very decent, and can be much better than that of some previous approaches like semi-DI protocols.

It should be pointed out that, when training neural networks, we only need to measure sampled quantum states locally, or even by single-qubit Pauli observables, and then collect the outcome statistics data as the corresponding features, which is actually quite experiment-friendly, as large-scale quantum measurements are difficult to implement physically, especially for high-dimensional quantum states. Meanwhile, we would like to point out that our numerical experiments show that sometimes the approach still enjoys comparable performance if one reduces the outcome numbers of involved quantum measurements for each party, which will make the size of features smaller, and thus improve the efficiency of both the collection of measurement outcome statistics data and the training of neural networks.

In summary, with the rise of quantum engineering in recent years, it is very realistic to consider applying neural networks widely in quantum engineering problems where a rough quantification of underlying entanglement is helpful. Because of this, it will be interesting to improve the performance of neural networks further by introducing new techniques and try to utilize neural networks to quantify more measures of quantum entanglement.

Acknowledgements This work was supported by the National Key R&D Program of China, Grant No. 2018YFA0306703, 2021YFE0113100, and the National Natural Science Foundation of China, Grant No. 61832015, 62272259.

Data availability The data that support the findings of this study are available upon reasonable request from the authors.

Declarations

Conflict of interest The authors declare no competing interests.

References

1. Paris, M., Rehacek, J.: Quantum State Estimation, vol. 649. Springer, New York (2004)
2. Steffen, M., Ansmann, M., Radoslaw, C.B., Nadav, K., Erik, L., Robert, M., Matthew, N., Eva, M.W., Andrew, N.C., John, M.M.: Measurement of the entanglement of two superconducting qubits via state tomography. Science **313**, 1423–1425 (2006)
3. Terhal, B.M.: Bell inequalities and the separability criterion. Phys. Lett. A **271**, 319–326 (2000)
4. Tura, J., Augusiak, R., Ana, B.S., Tamas, V., Maciej, L., Antonio, A.: Detecting nonlocality in many-body quantum states. Science **344**, 1256–1258 (2014)
5. Bavaresco, J., Natalia, H.V., Claude, K., Matej, P., Paul, E., Nicolai, F., Mehul, M., Marcus, H.: Measurements in two bases are sufficient for certifying high-dimensional entanglement. Nat. Phys. **14**, 1032–1037 (2018)
6. Rosset, D., Ferretti-Schöbitz, R., Bancal, J.-D., Gisin, N., Liang, Y.-C.: Imperfect measurement settings: implications for quantum state tomography and entanglement witnesses. Phys. Rev. A **86**, 062325 (2012)
7. Gühne, O., Tóth, G.: Entanglement detection. Phys. Rep. **474**, 1–75 (2009)

8. Friis, N., Vitagliano, G., Malik, M., Huber, M.: Entanglement certification from theory to experiment. Nat. Rev. Phys. **1**, 72–87 (2019)
9. Collins, D., Gisin, N., Popescu, S., Roberts, D., Scarani, V.: Bell-type inequalities to detect true n-body non separability. Phys. Rev. Lett. **88**, 170405 (2002)
10. Bancal, J.-D., Gisin, N., Liang, Y.-C., Pironio, S.: Device-independent witnesses of genuine multipartite entanglement. Phys. Rev. Lett. **106**, 250404 (2011)
11. Pál, K.F., Vértesi, T.: Multisetting bell-type inequalities for detecting genuine multipartite entanglement. Phys. Rev. A **83**, 062123 (2011)
12. Murta, G., Ramanathan, R., Móller, N., Marcelo, T.C.: Quantum bounds on multiplayer linear games and device-independent witness of genuine tripartite entanglement. Phys. Rev. A **93**, 022305 (2016)
13. Baccari, F., Cavalcanti, D., Wittek, P., Acín, A.: Efficient device-independent entanglement detection for multipartite systems. Phys. Rev. X **7**, 021042 (2017)
14. Tavakoli, A., Alastair, A.A., Marc-Olivier, R., Nicolas, G., Nicolas, B.: Semi-device-independent characterization of multipartite entanglement of states and measurements. Phys. Rev.A **98**, 052333 (2018)
15. Zwerger, M., Dür, W., Bancal, J.-D., Sekatski, P.: Device-independent detection of genuine multipartite entanglement for all pure states. Phys. Rev. Lett. **122**, 060502 (2019)
16. Horodecki, R., Horodecki, P., Horodecki, M., Horodecki, K.: Quantum entanglement. Rev. Mod. Phys. **81**, 865 (2009)
17. Eltschka, C., Siewert, J.: Quantifying entanglement resources. J. Phys. A Math. Theor. **47**, 424005 (2014)
18. Życzkowski, K., Horodecki, P., Sanpera, A., Lewenstein, M.: Volume of the set of separable states. Phys. Rev. A **58**, 883 (1998)
19. Eisert, J., Plenio, M.B.: A comparison of entanglement measures. J. Mod. Opt. **46**, 145–154 (1999)
20. Daley, A.J., Pichler, H., Schachenmayer, J., Zoller, P.: Measuring entanglement growth in quench dynamics of bosons in an optical lattice. Phys. Rev. Lett. **109**, 020505 (2012)
21. Abanin, D.A., Demler, E.: Measuring entanglement entropy of a generic many-body system with a quantum switch. Phys. Rev. Lett. **109**, 020504 (2012)
22. Brydges, T., Elben, A., Jurcevic, P., Vermersch, B., Maier, C., Ben, P.L., Peter, Z., Rainer, B., Christian, F.R.: Probing rényi entanglement entropy via randomized measurements. Science **364**, 260–263 (2019)
23. Zhou, Y., Zeng, P., Liu, Z.: Single-copies estimation of entanglement negativity. Phys. Rev. Lett. **125**, 200502 (2020)
24. Gray, J., Banchi, L., Bayat, A., Bose, S.: Machine-learning-assisted many-body entanglement measurement. Phys. Rev. Lett. **121**, 150503 (2018)
25. Berkovits, R.: Extracting many-particle entanglement entropy from observables using supervised machine learning. Phys. Rev. B **98**, 241411 (2018)
26. Shahandeh, F., Hall, M.J.W., Ralph, T.C.: Measurement-device-independent approach to entanglement measures. Phys. Rev. Lett. **118**, 150505 (2017)
27. Rosset, D., Martin, A., Verbanis, E., Lim, Charles C.W., Thew, R.: Practical measurement-device-independent entanglement quantification. Phys. Rev. A **98**, 052332 (2018)
28. Guo, Y., Bai-Chu, Y., Xiao-Min, H., Liu, B.-H., Yu-Chun, W., Huang, Y.-F., Li, C.-F., Guo, G.-C.: Measurement-device-independent quantification of irreducible high-dimensional entanglement. npj Quantum Inf. **6**, 1–6 (2020)
29. Moroder, T., Bancal, J.-D., Liang, Y.-C., Hofmann, M., Gühne, O.: Device-independent entanglement quantification and related applications. Phys. Rev. Lett. **111**, 030501 (2013)
30. Bardyn, C.-E., Timothy, C.H.L., Serge, M., Matthew, M., Valerio, S.: Device-independent state estimation based on bell's inequalities. Phys. Rev. A **80**, 062327 (2009)
31. Kaniewski, J.: Analytic and nearly optimal self-testing bounds for the clauser-horne-shimony-holt and mermin inequalities. Phys. Rev. Lett. **117**, 070402 (2016)
32. Wei, Z., Lin, L.: Analytic semi-device-independent entanglement quantification for bipartite quantum states. Phys. Rev. A **103**, 032215 (2021)
33. Horodecki, M., Horodecki, P., Horodecki, R.: Limits for entanglement measures. Phys. Rev. Lett. **84**, 2014 (2000)
34. Lijinzhi, L., Zhaohui W.: Quantifying multipartite quantum entanglement by nondegenerate bell inequalities. arXiv preprint arXiv:2008.12064 (2020)

35. Elben, A., Kueng, R., Huang, H.-Y.R., van Bijnen, R., Kokail, C., Dalmonte, M., Calabrese, P., Kraus, B., Preskill, J., Zoller, P., et al.: Mixed-state entanglement from local randomized measurements. Phys. Rev. Lett. **125**, 200501 (2020)
36. Devetak, I., Winter, A.: Distillation of secret key and entanglement from quantum states. Proc. R. Soc. A Math. Phys. Eng. Sci. **461**, 207–235 (2005)
37. Schumacher, B., Nielsen, M.A.: Quantum data processing and error correction. Phys. Rev. A **54**, 2629 (1996)
38. Lloyd, S.: Capacity of the noisy quantum channel. Phys. Rev. A **55**, 1613 (1997)
39. Shimony, A.: Degree of entanglement a. Ann. N. Y. Acad. Sci. **755**, 675–679 (1995)
40. Barnum, H., Linden, N.: Monotones and invariants for multi-particle quantum states. J. Phys. A Math. Gen. **34**, 6787 (2001)
41. Collins, D., Gisin, N., Linden, N., Massar, S., Popescu, S.: Bell inequalities for arbitrarily high-dimensional systems. Phys. Rev. Lett. **88**, 040404 (2002)
42. Acin, A., Durt, T., Gisin, N., José, I.L.: Quantum nonlocality in two three-level systems. Phys. Rev. A **65**, 052325 (2002)
43. Chen, J.-L., Chunfeng, W., Leong, C.K., Choo, H.O., Mo-Lin, G.: Violating bell inequalities maximally for two d-dimensional systems. Phys. Rev. A **74**, 032106 (2006)
44. Zohren, S., Gill, R.D.: Maximal violation of the collins-gisin-linden-massar-popescu inequality for infinite dimensional states. Phys. Rev. Lett. **100**, 120406 (2008)
45. Lin, X., Chen, Z., Wei, Z.: Quantifying quantum entanglement via a hybrid quantum-classical machine learning framework. Phys. Rev. A **107**, 062409 (2023)
46. Barrett, J., Kent, A., Pironio, S.: Maximally nonlocal and monogamous quantum correlations. Phys. Rev. Lett. **97**, 170409 (2006)
47. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proc. IEEE **86**, 2278–2324 (1998)
48. Albawi, S., Mohammed, T. A., Al-Zawi, S.: Understanding of a convolutional neural network. In: 2017 International Conference on Engineering and Technology (ICET), pp. 1–6. IEEE (2017)
49. Raschka, S.: Model evaluation, model selection, and algorithm selection in machine learning. arXiv preprint arXiv:1811.12808 (2018)
50. Efron, B., Tibshirani, R.J.: An introduction to the bootstrap. CRC Press, Florida (1994)
51. Byrd, R.H., Peihuang, L., Nocedal, J., Zhu, C.: A limited memory algorithm for bound constrained optimization. SIAM J. Sci. Comput. **16**, 1190–1208 (1995)
52. Zhu, C., Richard, H.B., Peihuang, L., Jorge, N.: Algorithm 778: L-bfgs-b: Fortran subroutines for large-scale bound-constrained optimization. ACM Trans. Math. Softw. (TOMS) **23**, 550–560 (1997)
53. Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., et al.: Scipy 1.0: fundamental algorithms for scientific computing in python. Nat. Methods **17**, 261–272 (2020)
54. Dür, W., Vidal, G.: Three qubits can be entangled in two inequivalent ways. Phys. Rev. A **62**, 062314 (2000)
55. Daniel, M.G., Michael, A.H., Anton, Z.: Going Beyond Bell's Theorem, in Bell's Theorem, Quantum Theory and Conceptions of the Universe, pp. 69–72. Springer, New York pp (1989)
56. Dominic M., Andrew Y: Self testing quantum apparatus. arXiv preprint arXiv:quant-ph/0307205 (2003)
57. Šupić, I., Bowles, J.: Self-testing of quantum systems: a review. Quantum **4**, 337 (2020)