Machine Learning Nonlocal Correlations

Askery Canabarro,[1,2] Samuraí Brito,[1] and Rafael Chaves[1,3]
[1]International Institute of Physics, Federal University of Rio Grande do Norte, 59070-405 Natal, Brazil
[2]Grupo de Física da Matéria Condensada, Núcleo de Ciências Exatas—NCEx, Campus Arapiraca, Universidade Federal de Alagoas, 57309-005 Arapiraca-AL, Brazil
[3]School of Science and Technology, Federal University of Rio Grande do Norte, 59078-970 Natal, Brazil

(Received 24 August 2018; published 22 May 2019)

The ability to witness nonlocal correlations lies at the core of foundational aspects of quantum mechanics and its application in the processing of information. Commonly, this is achieved via the violation of Bell inequalities. Unfortunately, however, their systematic derivation quickly becomes unfeasible as the scenario of interest grows in complexity. To cope with that, here, we propose a machine learning approach for the detection and quantification of nonlocality. It consists of an ensemble of multilayer perceptrons blended with genetic algorithms achieving a high performance in a number of relevant Bell scenarios. As we show, not only can the machine learn to quantify nonlocality, but discover new kinds of nonlocal correlations inaccessible with other current methods as well. We also apply our framework to distinguish between classical, quantum, and even postquantum correlations. Our results offer a novel method and a proof-of-principle for the relevance of machine learning for understanding nonlocality.

DOI: 10.1103/PhysRevLett.122.200401

Quantum correlations, stronger than those allowed by classical systems, are at the core of quantum information science, its fundamental implications, and practical applications [1,2]. For instance, the correlations obtained by measurements on distant entangled particles are incompatible with any local hidden variable (LHV) model [3], a cornerstone in our understanding of quantum theory that paved the way to many relevant information processing tasks ranging from quantum cryptography [4–6] and randomness certification [7,8] to self-testing [9] and distributed computing [10]. To that aim, it is crucial to develop ways to test the incompatibility of a given correlation with LHV models, that is, to detect its nonlocal behavior.

The most common approach to that purpose is based on Bell inequalities. First, their violation is an unambiguous witness of the nonclassicality of the correlations. Second, they serve as an objective function over which one can optimize quantum states and measurements to find violations and, thus, search for nonlocal correlations. Given its clear importance, over the years a very general framework has been developed [11,12], and dozens of inequalities were found [2]. LHV models define a set of correlations compatible with it, the nontrivial boundaries of which are precisely the Bell inequalities. Typically, however, the characterization of the local set via Bell inequalities is computationally very demanding, rapidly becoming intractable as the scenario of interest raises its complexity [13–16]. The situation is far worse for more general situations, for instance, when dealing with quantum networks [17–24] where many independent sources of entangled states are present and give rise to semialgebraic (nonconvex) sets [25,26]. Faced with this impairing situation, it is natural to search for alternative routes that do not rely on Bell inequalities. That is precisely the aim of this work.

Motivated by the outstanding recent progress within quantum physics [27–34], here, we propose a machine learning (ML) approach to test the nonclassicality of correlations. Our starting point is a recently introduced quantifier of nonlocality [35] considering "how far" a given correlation is from the local set. As opposed to a specific Bell inequality—covering a very limited region of the space of correlations—our approach offers a global perspective of the local set geometry, in some sense testing all Bell inequalities at once. We randomly sample the space of correlations, compute this quantifier, and feed this data to an ensemble of deep learning algorithms [36] able to recognize patterns in the correlations and create machine models for the complex geometry of Bell correlations.

Strikingly, not only can the neural networks quantify, with a high accuracy, the nonlocality of given correlations, but can also be used to discover new nonlocal correlations that could hardly be found by any other available means. That is, not only can the machine learn, but also teach us something new about Bell correlations. This is shown in a variety of Bell scenarios, including the simplest scenario for which no complete characterization of the local set (Bell inequalities) and the entanglement swapping experiment [37–40] giving rise to the notoriously thorny bilocality scenario [17–19]. Finally, we also show that the machine can distinguish between quantum and postquantum correlations [41,42], an important topic in the foundations of quantum theory [43].

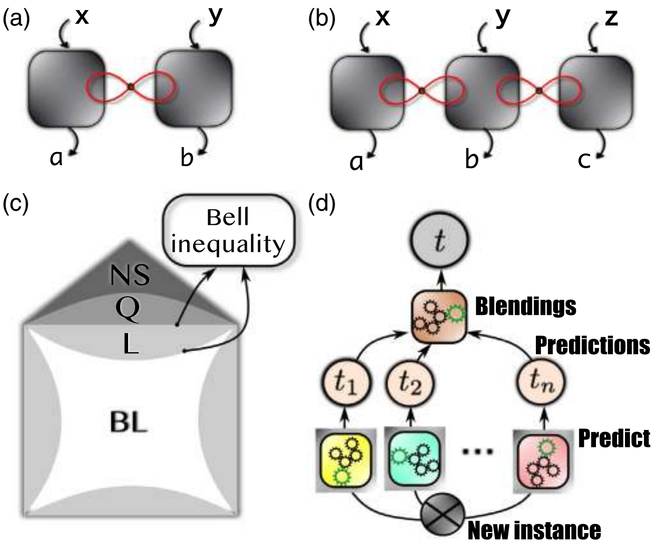

FIG. 1. Black-box representation of (a) the bipartite Bell scenario and (b) a tripartite scenario with two independent sources of states. (c) Pictorial illustration of the different sets of correlations: nonsignaling, quantum, local, and bilocal. (d) Blending technique where different machines are combined to improve the overall performance.

A machine learning approach to detect and quantify nonlocal correlations.—Bell's theorem [3] shows that measurements on distant entangled systems are incompatible with the assumption of local realism. We will refer to the simplest Bell scenario [see Fig. 1(a)], composed of two distant parties that, upon receiving their shares of a composite physical system, measure different observables (labeled by the variables X and Y) obtaining a respective measurement outcome (labeled by A and B). In a classical description, the probability distribution $p(A=a, B=b|X=x, Y=y) = p(a,b|x,y)$ observed in such a simple experiment should be decomposable in terms of a LHV model, that is,

$$p(a,b|x,y) = \sum_{\lambda} p(a|x,\lambda)p(b|y,\lambda)p(\lambda), \qquad (1)$$

defining a convex set $\mathcal{L}$, the boundaries of which are Bell inequalities [see Fig. 1(c)]. According to Born's rule, however, quantum mechanics implies that

$$p(a,b|x,y) = \mathrm{Tr}[(M_a^x \otimes M_b^y)\varrho_{AB}], \qquad (2)$$

where ϱ_{AB} is the density operator describing the shared physical system, and M_a^x and M_b^y describe measurement operators. Thus, to test the nonlocality of a given quantum distribution (2), we have to show that it falls outside the set $\mathcal{L}$, the paradigmatic method for that being the violation of a Bell inequality.

However, the number of Bell inequalities grows very fast as the Bell scenario of interest grows its complexity (number of parties, measurements, or outcomes) [2,12,13], that is, any given inequality will typically offer very limited and localized information of a high-dimensional and intricate set of correlations. To cope with that, we employ a more refined description here, based on a nonlocality quantifier $\mathrm{NL}(\mathbf{q})$ given by minimum trace distance between the distribution $\mathbf{q} = q(a,b|x,y)$ under test and a $\mathbf{p} = p(a,b|x,y)$ in the set of local distributions [35]

$$\mathrm{NL}(\mathbf{q}) = \frac{1}{2|x||y|} \min_{\mathbf{p}\in\mathcal{L}} \sum_{a,b,x,y} |\mathbf{q} - \mathbf{p}|, \qquad (3)$$

where $|x| = |y| = m$ denotes the number of possible measurements performed by the parties.

Having defined a Bell scenario of interest, the first step in our ML approach is to generate the training points to the machine. We do that by randomly sampling nonsignaling (NS) distributions defined by simple linear constraints (see the Supplemental Material (SM) [44]). The reason for sampling NS instead of quantum distributions is threefold. First, characterizing the quantum set is extremely challenging, the best available method given by an infinite hierarchy of semidefinite programs [42]. Second, even thought the NS condition allows for correlations beyond quantum mechanics, they play an important role in the foundations of the theory [41,43]. Third, the NS machine model describes quantum correlations equally well. For each sampled correlation [45], we compute the corresponding exact measure $\mathrm{NL}_{\mathrm{test}}$ (the target function t), feeding this data to different neural networks, the best ones have been blended via a genetic algorithm [46] to generate a machine prediction $\mathrm{NL}_{\mathrm{predicted}}$ [see Fig. 1(d) and SM [44] for details]. To measure the performance/error of the model, we employ the average trace distance $P = (1/N)\sum_{i=1}^{N} |\mathrm{NL}_{\mathrm{test}}^i - \mathrm{NL}_{\mathrm{predicted}}^i|$, where N is the number of points in the testing set [47].

Our first goal is to show that the machine can learn in a reliable manner to detect and quantify nonlocality. The results for the bipartite Bell scenario with $m = 2$, 3, 4, 5 dichotomic measurements and input data of 5×10^5 points are shown in Fig. 2 and Table I. The average error is of order 10^{-3} in all scenarios. The measure $\mathrm{NL}(\mathbf{q})$ is a function of all Bell inequalities defining a given scenario, and its number is equal to $8(m=2)$, $72(m=3)$, $27936(m=4)$ while already, for $m=5$, no complete characterization is available. Thus, such high accuracies are a truly striking feature of the deep learning approach. But what can we learn from the machine?

First, we notice that, once the machine model is trained, to obtain a prediction about a new instance is basically instantaneous, in some cases offering a 10^5 speedup (more on that below). Second, in spite of the machine being trained over the NS set (including postquantum correlations), it provides a remarkably accurate description of the quantum set as well [see in Fig. 2(a)]. Finally, the machine model can be used to unveil new kinds of nonlocal

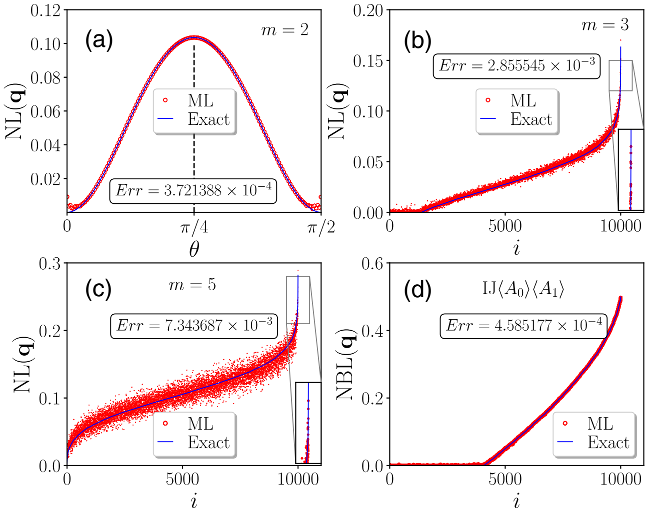

FIG. 2. The straight line (blue) indicates the exact solution of Eq. (3) or Eq. (6) (see SM [44]) and the circle (red) indicates the ML prediction (considering 10^4 test set points). In all cases, the machine can predict, with excellent accuracy, the degree of nonlocality without any information about Bell inequalities. (a) Bipartite scenario ($m = 2$) with quantum correlations obtained by projective measurements on $|\psi\rangle = \cos\theta|00\rangle + \sin\theta|11\rangle$, that maximally violate the Clauser-Horne-Shimony-Holt inequality [48]. (b) Bipartite scenario ($m = 3$), (c) ($m = 5$), and (d) the bilocality scenario employing four features $(I, J, \langle A_0\rangle, \langle A_1\rangle)$. For (b), (c), and (d), the x axis denotes the point ith correlations ($\mathbf{q}_i$) with an increasing ordering in the value of $\mathrm{NL}^i_{\mathrm{test}}$.

correlations that would be unaccessible by standard approaches. Consider, for instance, a typical situation where the experimental platform imposes strict constraints on the types of measurements that can be performed and their efficiency. The paradigmatic approach would be to optimize (under specified constraints) the violation of some Bell inequality. However, as said before, very few classes of inequalities are known, and often, they provide extremely limited information. In contrast, the ML approach provides an objective function taking the whole local set into account that can be optimized over to generate candidate nonlocal correlations. We say candidate correlations because, as in any other ML construction, there will be an associated intrinsic uncertainty (however small). Nonetheless, once a good candidate is found, one can use the linear programming approach (generating the data training the machine) to certify its nonclassicality. A specific example of how we can learn from the machine is given below, considering the thorny bilocality scenario.

Nonlocal correlations in a simple quantum network.—To illustrate the power of the ML approach, consider the simplest possible quantum network beyond Bell's paradigmatic scenario, akin to an entanglement swapping experiment [37]. It consists of three separated parties interconnected by two independent sources of states [see Fig. 1(b)]. The LHV model, taking into account the independence of the sources—$p(\lambda_1, \lambda_2) = p(\lambda_1)p(\lambda_2)$, the bilocality assumption [17,18]—implies that the tripartite observed distribution can be written as

$$p(a, b, c|x, y, z) = \sum_{\lambda_1,\lambda_2} p(a|x, \lambda_1)p(b|y, \lambda_1, \lambda_2)p(c|z, \lambda_2)p(\lambda_1)p(\lambda_2). \quad (4)$$

Interestingly, there are local correlations that, nonetheless, are nonbilocal (NBL); correlations that appear of classical nature have their nonclassicality revealed if the independence of the sources generating the correlations is considered [see Fig. 1(c)]. On the negative side, Eq. (4) defines an intricate nonconvex set for which very few and specific inequalities have been derived [17–23]. The ML approach offers a novel way to circumvent this difficulty. Similar to what has been done before, we sample over NS distributions and compute its distance $\mathrm{NBL}(\mathbf{q})$ to the bilocal set via a sequence of linear programs (see [49] and SM [44] for details).

Consider that the three parties perform two possible measurements ($x, y, z = 0, 1$). The only known inequality in this scenario is given by

$$\sqrt{|I|} + \sqrt{|J|} \leq 1, \quad (5)$$

with $I = (1/4)\sum_{x,z}\langle A_x, B_0, C_z\rangle$ and $J = (1/4)\sum_{x,z}(-1)^{x+z}\times\langle A_x, B_1, C_z\rangle$. Within that context, we considered two different cases, both considering a total of 1.5×10^5 sampled distributions. In the first, the correlations are encoded in the expectation values $(I, J, \langle A_0\rangle, \langle A_1\rangle)$ (four features) and in the second $(\langle A_0B_0, C_0\rangle, \ldots, \langle A_1B_1, C_1\rangle, \langle A_0\rangle, \langle A_1\rangle)$ (ten features) [50]. The results are shown in Fig. 3 and in Table I. The overall performance is very high.

We have also compared the ML models trained with NS correlations to detect quantum ones. On the first,

TABLE I. Performance (average ℓ_1-norm) for different scenarios and different ML approaches. See SM [44] for details.

	Scenario					
Technique	$m = 2$	$m = 3$	$m = 4$	$m = 5$	IJA_0A_1	$A_xB_yC_z, A_x$
Typical MLP $(\times 10^{-3})$	0.46	2.20	7.75	8.50	2.70	6.30
Blending $(\times 10^{-3})$	0.05	1.54	6.78	7.31	0.45	3.22

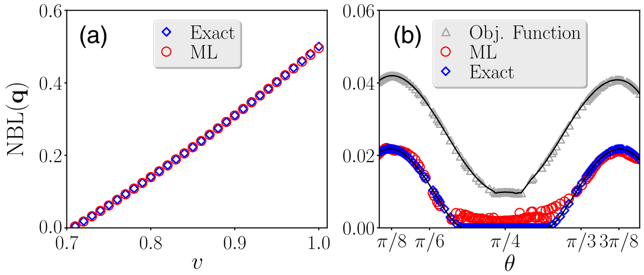

FIG. 3. Bilocality scenario. (a) Diamonds (blue), indicate the exact value of $\mathrm{NBL} = v^2 - 1/2$ [49] obtained by measurements on a Werner state maximally violating the inequality (5). The circles (red) indicate the deep learning prediction. (b) The triangles (grey) indicate the results of a numerical optimization for the maximum value of the regression ML function for quantum correlations obtained by measurements on the state $|\psi\rangle_{AB} = |\psi\rangle_{BC} = \cos\theta|00\rangle + \sin\theta|11\rangle$ (not violating inequality 5). The diamonds (blue) indicate the exact value and the circles (red) indicate the prediction made by a neural network trained with NS correlations. Strikingly, the ML approach can discover new quantum correlations without any information about Bell inequalities.

we have used the blended ensemble of deep learning models to compute the degree of nonbilocality of the correlations obtained by measurements on a Werner state $\varrho = v|\Phi^+\rangle\langle\Phi^+| + (1-v)\mathbb{1}/4$ [with $|\Phi^+\rangle = (1/\sqrt{2})(|00\rangle + |11\rangle)$] and that maximally violate the inequality (5), obtaining nearly perfect agreement [see Fig. 3(a)]. On the second, we have numerically searched for quantum correlations violating the ML regression function but that do not violate (5). As discussed above, in this case, our ML approach is providing us with new and relevant information: the machine provides us new examples of correlations, the nonclassicality of which cannot be detected by the only known inequality (5) [see Fig. 3(b)]. We highlight that, after training the machine, to obtain a prediction about a new instance is basically instantaneous (of the order of 10^{-4} seconds) while the brute force method (used to train the machine) takes considerably more time, on average 20 seconds, thus, offering a 10^5 speedup.

Machine learning postquantum correlations.—The best available method to characterize the set of quantum correlations (those obtainable by measurements on a quantum state) is given by a hierarchy of semidefinite programs that converges asymptotically to the quantum set [42] and, thus, only provides an outer approximation, in general. Notwithstanding, in some particular instances, the convergence happens at a finite step, as is the case in a bipartite scenario where each party can perform two possible dichotomic measurements.

A necessary and sufficient condition [51] for the expectation values $\langle A_x B_y\rangle$ with $x, y = 0, 1$ to have a quantum realization is given by all four symmetries of the inequality $|\arcsin\langle A_0B_0\rangle + \arcsin\langle A_0B_1\rangle + \arcsin\langle A_1B_0\rangle - \arcsin\langle A_1B_1\rangle| \leq \pi$. Furthermore, the nonlocality of the associated distribution can also be decided by testing all the symmetries of the inequality $|\langle A_0B_0\rangle + \langle A_0B_1\rangle + \langle A_1B_0\rangle - \langle A_1B_1\rangle| \leq 2$. Given the list of correlators $\langle A_x B_y\rangle$, we can then classify it as local, nonlocal (quantum), or postquantum.

TABLE II. The confusion matrix C_{ij} of the blend of classifiers for 10^5 unseen inputs, which returns the number of observations known to be in group i but predicted to be in group j. The sum of the elements of the main diagonal divided by the total of elements gives the accuracy score.

	Predictions		
True Class	Local	Quantum	Postquantum
Local	33436	96	0
Quantum	41	33173	236
Postquantum	0	136	32882

In machine learning, classification is the problem of determining to which class of categories a new observation belongs by means of a training set of data containing instances whose category membership is known [52]. The ensemble of classifiers, created in a similar way as the ensemble of regressors, was trained over 4×10^5 input points. The overall accuracy achieved was 99.49%. To better quantify the quality of the predictions of the ensemble of deep learning models that we proposed, we computed the confusion matrix for a random sample of 10^5 unseen new instances in Table II, see SM [44] for more details. Interestingly, even though most of the postquantum points occur close to the local set [53], the ML method never makes mistakes between both. That is, in spite of their very small geometrical distance, the machine perceives traits in those correlations allowing us to distinguish between them.

Discussion.—Bell nonlocality shows that, even without a precise description of a physical apparatus and solely based on measurement data, one can prove the quantumness of some observed correlations. It is at the core of the device-independent approach to quantum information processing [54] with many applications in near term quantum technologies such as quantum cryptography [4–6]. Detecting Bell nonlocality beyond simple cases, however, remains a thorny issue given the hard computational complexity of the characterization of locality via Bell inequalities [12,13]. Further, with the recent advances on the quantum internet [55,56]—in short, a network with several independent sources of quantum states—such computational difficulties become even more pronounced [19,25,26]. Here, we propose an alternative and timely route, a machine learning approach, allowing the detection and quantification of nonlocality as well as its quantum (or postquantum) nature. To illustrate its benefits, we have applied it to a number of relevant Bell scenarios showing that not only can the machine learn, but also teach, for instance, pointing to new kinds of nonlocal correlations that cannot be detected by known Bell inequalities. From the machine learning side, to the best of our knowledge, it is the first time that many multilayer perceptrons (MLPs) are blended using generative algorithms to automatically find the best way to combine them, proving a robust framework for dealing with a great variety of Bell scenarios.

Our results provide a proof-of-principle for the relevance of ML tools in Bell nonlocality, and we trust they will open several research venues. A natural next step is to consider classical and quantum networks of growing complexity [19–24]. A clear bottleneck of our approach is the fact that we have to train the machine and this becomes very computationally costly as we increase the complexity. On the positive side, once the ML model is obtained, it provides an enormous speedup and our framework can naturally be integrated with more efficient training methods [57,58] alternative to linear programming. Furthermore, to investigate alternative ML approaches, such as anomaly detection, is certainly interesting. In that case, the machine should be trained with local correlations only, clearly a much simpler training task. Another clear possibility is the combination with other recent results, e.g., the reinforcement learning approach to finding the maximum violation of a given Bell inequality [34]. Finally, one can wonder how much we can learn from the machine models. For instance, in spite of their very close geometrical distance, the machine almost perfectly recognizes classical, quantum, and postquantum correlations. Can it be that the machine is recognizing some new physical principle presently unknown? If so, how can we retrieve it? That means we have to somehow open the machine black box [59], and we hope our results motivate further research in this direction.

The authors acknowledge the Brazilian ministries MEC and MCTIC, funding agency CNPq (A. C.'s Universal Grants No. 423713/2016-7, No. 307172/2017-1 and No. 406574/2018-9 and INCT-IQ) and UFAL (A. C.'s paid license for scientific cooperation at UFRN). This work was supported by the Serrapilheira Institute (Grant No. Serra-1708-15763) and the John Templeton Foundation via Grant Q-CAUSAL No. 61084.

[1] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, Quantum entanglement, Rev. Mod. Phys. **81**, 865 (2009).
[2] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner, Bell nonlocality, Rev. Mod. Phys. **86**, 419 (2014).
[3] J. S. Bell, On the Einstein-Podolsky-Rosen paradox, Physics **1**, 195 (1964).
[4] A. K. Ekert, Quantum Cryptography Based on Bell's Theorem, Phys. Rev. Lett. **67**, 661 (1991).
[5] J. Barrett, L. Hardy, and A. Kent, No Signaling and Quantum Key Distribution, Phys. Rev. Lett. **95**, 010503 (2005).
[6] A. Acín, N. Brunner, N. Gisin, S. Massar, S. Pironio, and V. Scarani, Device-Independent Security of Quantum Cryptography Against Collective Attacks, Phys. Rev. Lett. **98**, 230501 (2007).
[7] R. Colbeck, Ph.D. thesis, University of Cambridge, 2007.
[8] S. Pironio *et al.*, Random numbers certified by Bell's theorem, Nature (London) **464**, 1021 (2010).
[9] D. Mayers and A. Yao, Self testing quantum apparatus, Quantum Inf. Comput. **4**, 273 (2004).
[10] H. Buhrman, R. Cleve, S. Massar, and R. de Wolf, Nonlocality and communication complexity, Rev. Mod. Phys. **82**, 665 (2010).
[11] A. Fine, Hidden Variables, Joint Probability, and the Bell Inequalities, Phys. Rev. Lett. **48**, 291 (1982).
[12] I. Pitowsky, *Quantum Probability Quantum Logic* (Springer-Verlag, Berlin, 1989).
[13] I. Pitowsky, Correlation polytopes: Their geometry and complexity, Math. Program. **50**, 395 (1991).
[14] D. Collins, N. Gisin, N. Linden, S. Massar, and S. Popescu, Bell Inequalities for Arbitrarily High-Dimensional Systems, Phys. Rev. Lett. **88**, 040404 (2002).
[15] D. Collins and N. Gisin, A relevant two qubit Bell inequality inequivalent to the CHSH inequality, J. Phys. A **37**, 1775 (2004).
[16] N. Brunner and N. Gisin, Partial list of bipartite Bell inequalities with four binary settings, Phys. Lett. A **372**, 3162 (2008).
[17] C. Branciard, N. Gisin, and S. Pironio, Characterizing the Nonlocal Correlations Created via Entanglement Swapping, Phys. Rev. Lett. **104**, 170401 (2010).
[18] C. Branciard, D. Rosset, N. Gisin, and S. Pironio, Bilocal versus nonbilocal correlations in entanglement-swapping experiments, Phys. Rev. A **85**, 032119 (2012).
[19] R. Chaves, Polynomial Bell Inequalities, Phys. Rev. Lett. **116**, 010402 (2016).
[20] D. Rosset, C. Branciard, T. Jack Barnea, G. Pütz, N. Brunner, and N. Gisin, Nonlinear Bell Inequalities Tailored for Quantum Networks, Phys. Rev. Lett. **116**, 010403 (2016).
[21] A. Tavakoli, Bell-type inequalities for arbitrary noncyclic networks, Phys. Rev. A **93**, 030101(R) (2016).
[22] C. M. Lee and M. J. Hoban, Towards Device-Independent Information Processing on General Quantum Networks, Phys. Rev. Lett. **120**, 020504 (2018).
[23] M.-X. Luo, Computationally Efficient Nonlinear Bell Inequalities for Quantum Networks, Phys. Rev. Lett. **120**, 140402 (2018).
[24] A. Kela, K. von Prillwitz, J. Aberg, R. Chaves, and D. Gross, Semidefinite tests for latent causal structures, arXiv:1701.00652.
[25] D. Geiger and C. Meek, Quantifier elimination for statistical problems, in *Uncertainty in Artificial Intelligence: Proceedings of the 15th Conference, Royal Institute of Technology (KTH), Stockholm, Sweden, 1999* (Morgan Kaufmann Publishers Inc., Stockholm, 1999), p. 226.
[26] L. D. Garcia, M. Stillman, and B. Sturmfels, Algebraic geometry of Bayesian networks, J. Symb. Comput. **39**, 331 (2005).
[27] J. Carrasquilla and R. G. Melko, Machine learning phases of matter, Nat. Phys. **13**, 431 (2017).

[28] K. Ch'ng, J. Carrasquilla, R. G. Melko, and E. Khatami, Machine Learning Phases of Strongly Correlated Fermions, Phys. Rev. X **7**, 031038 (2017).

[29] P. Mehta, M. Bukov, C.-H. Wang, A. G. R. Day, C. Richardson, C. K. Fisher, and D. J. Schwab, A high-bias, low-variance introduction to machine learning for physicists, arXiv:1803.08823.

[30] G. Torlai, G. Mazzola, J. Carrasquilla, M. Troyer, R. Melko, and G. Carleo, Neural-network quantum state tomography, Nat. Phys. **14**, 447 (2018).

[31] X. Gao and L.-M. Duan, Efficient representation of quantum many-body states with deep neural networks, Nat. Commun. **8**, 662 (2017).

[32] G. Carleo and M. Troyer, Solving the quantum many-body problem with artificial neural networks, Science **355**, 602 (2017).

[33] Y.-C. Ma and M.-H. Yung, Transforming Bell's inequalities into state classifiers with machine learning, arXiv:1705.00813.

[34] D.-L. Deng, Machine Learning Detection of Bell Nonlocality in Quantum Many-Body Systems, Phys. Rev. Lett. **120**, 240402 (2018).

[35] S. G. A. Brito, B. Amaral, and R. Chaves, Quantifying Bell nonlocality with the trace distance, Phys. Rev. A **97**, 022111 (2018).

[36] J. Schmidhuber, Deep learning in neural networks: An overview, Neural Netw. **61**, 85 (2015).

[37] M. Zukowski, A. Zeilinger, M. A. Horne, and A. K. Ekert, Event-Ready-Detectors Bell Experiment via Entanglement Swapping, Phys. Rev. Lett. **71**, 4287 (1993).

[38] G. Carvacho, F. Andreoli, L. Santodonato, M. Bentivegna, R. Chaves, and F. Sciarrino, Experimental violation of local causality in a quantum network, Nat. Commun. **8**, 14775 (2017).

[39] D. J. Saunders, A. J. Bennet, C. Branciard, and G. J. Pryde, Experimental demonstration of nonbilocal quantum correlations, Sci. Adv. **3**, e1602743 (2017).

[40] F. Andreoli, G. Carvacho, L. Santodonato, M. Bentivegna, R. Chaves, and F. Sciarrino, Experimental bilocality violation without shared reference frames, Phys. Rev. A **95**, 062315 (2017).

[41] S. Popescu and D. Rohrlich, Quantum nonlocality as an axiom, Found. Phys. **24**, 379 (1994).

[42] M. Navascués, S. Pironio, and A. Acín, Bounding the Set of Quantum Correlations, Phys. Rev. Lett. **98**, 010401 (2007).

[43] S. Popescu, Nonlocality beyond quantum mechanics, Nat. Phys. **10**, 264 (2014).

[44] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.122.200401 for details about the machine learning techniques and linear program formulation for the computation of $\mathrm{NL}(\mathbf{q})$ and $\mathrm{NBL}(\mathbf{q})$.

[45] To simplify the problem, without loss of generality, we do not use the full distribution as the input but, rather, the bipartite expectation value $\langle A_x B_y \rangle$ (see Appendix for general case). The measure $\mathrm{NL}(\mathbf{q})$ is optimized over all $q(a,b|x,y)$ compatible with a given $\langle A_x B_y \rangle$.

[46] R. S. Olson and J. H. Moore, Tpot: A tree-based pipeline optimization tool for automating machine learning, in *Proceedings of the Workshop on Automatic Machine Learning*, Proceedings of Machine Learning Research Vol. 64 (PMLR, 2016), pp. 66–74.

[47] Following the standard approach, the data are split in a training and cross validation (75%) and test (25%) dataset, the first and second used to create a machine model generating a prediction $\mathrm{NL}_{\mathrm{predicted}}$ and the second to test its accuracy in relation to the test targets $\mathrm{NL}_{\mathrm{test}}$.

[48] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, Proposed Experiment to Test Local Hidden-Variable Theories, Phys. Rev. Lett. **23**, 880 (1969).

[49] R. Chaves, R. Kueng, J. B. Brask, and D. Gross, Unifying Framework for Relaxations of the Causal Assumptions in Bell's Theorem, Phys. Rev. Lett. **114**, 140403 (2015).

[50] The marginal information provided by $\langle A_0 \rangle$ is essential to compute $\mathrm{NBL}(\mathbf{q})$. Please see Appendix for more details.

[51] L. Masanes, Necessary and sufficient condition for quantum-generated correlations, arXiv:quant-ph/0309137.

[52] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning* (MIT Press, 2016), http://www.deeplearningbook.org.

[53] C. Duarte, S. Brito, B. Amaral, and R. Chaves, Concentration phenomena in the geometry of Bell correlations, Phys. Rev. A **98**, 062114 (2018).

[54] S. Pironio, V. Scarani, and T. Vidick, Focus on device independent quantum information, New J. Phys. **18**, 100202 (2016).

[55] H. Jeff Kimble, The quantum internet, Nature (London) **453**, 1023 (2008).

[56] D. Castelvecchi, The quantum internet has arrived (and it hasn't), Nature (London) **554**, 289 (2018).

[57] A. Montina and S. Wolf, Can non-local correlations be discriminated in polynomial time? arXiv:1609.06269.

[58] S. Brierley, M. Navascues, and T. Vertesi, Convex separation from convex optimization for large-scale problems, arXiv:1609.05011.

[59] N. Walchover, New theory cracks open the black box of deep learning, Quanta Mag. **3** (2017).