Data-driven criteria for quantum correlations

Mateusz Krawczyk,[1] Jarosław Pawłowski,[1] Maciej M. Maśka,[1] and Katarzyna Roszak[2]

[1]*Institute of Theoretical Physics, Wrocław University of Science and Technology, Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland*
[2]*Institute of Physics (FZU), Czech Academy of Sciences, Na Slovance 2, 182 00 Prague, Czech Republic*

(Received 4 August 2023; accepted 9 January 2024; published 5 February 2024)

We build a machine learning model to detect correlations in a three-qubit system using a neural network trained in an unsupervised manner on randomly generated states. The network is forced to recognize separable states, and correlated states are detected as anomalies. Quite surprisingly, we find that the proposed detector performs much better at distinguishing a weaker form of quantum correlations, namely, the quantum discord, than entanglement. In fact, it has a tendency to grossly overestimate the set of entangled states even at the optimal threshold for entanglement detection, while it underestimates the set of discordant states to a much lesser extent. In order to illustrate the nature of states classified as quantum correlated, we construct a diagram containing various types of states—entangled, as well as separable, both discordant and nondiscordant. We find that the near-zero value of the recognition loss reproduces the shape of the nondiscordant separable states with high accuracy, especially considering the nontrivial shape of this set on the diagram. The network architecture is designed carefully: it preserves separability, and its output is equivariant with respect to qubit permutations. We show that the choice of architecture is important to get the highest detection accuracy, much better than for a baseline model that just utilizes a partial trace operation.

DOI: 10.1103/PhysRevA.109.022405

I. INTRODUCTION

Quantum correlations play a pivotal role in near-future quantum technologies, from quantum cryptography [1,2], to networks [3,4], to computation [5,6]. The source of their usefulness is that they allow for the state of one subsystem to be influenced by operations and measurements on the other, which yields effects that cannot be reproduced by classical physics. A good example here is teleportation [7–10], which allows for the transfer of an unknown quantum state between distant locations by means of local operations and measurements, and the transfer of classical information, due to a maximally entangled state being shared by two parties.

In quantum information theory, there is a long-standing question about quantum correlations in mixed states [11–13]. For pure states, any correlations between subsystems are quantum in their nature (entanglement), and the set of separable (not entangled) states is equivalent to the set of product states. In the case of mixed states, the fact that a density matrix cannot be written in product form does not guarantee the existence of quantum correlations, since this formalism allows for the description of classical (purely statistical) correlations, i.e., $\hat{\rho}_{AB} = 1/2(|00\rangle\langle 00| + |11\rangle\langle 11|)$, whereas a product state contains no correlations whatsoever.

One definition of the set of states that contain no quantum correlations, namely the set of mixed separable states, is given by [14–16]

$$\hat{\rho}_{AB}^{\text{sep}} = \sum_i p_i |\psi_i\rangle_{AA}\langle\psi_i| \otimes |\phi_i\rangle_{BB}\langle\phi_i|, \tag{1}$$

where p_i are probabilities and the indices A and B label the subsystems. It is relevant to note here that there is no limitation on states $|\psi_i\rangle_A$ or $|\phi_i\rangle_B$, which do not have to form a basis. By this definition, a bipartite state is separable if and only if it can be created from a product state using only local operations and classical communication. The quantification of entanglement (for mixed states, entangled states are defined as not separable) is classed as an NP-complete problem [17], and it does not simplify even when a state only needs to be qualified as separable or entangled.

A state with no quantum correlations can also be defined as a state for which no information about one subsystem can be learned by measurements performed on the other subsystem. Such states are said to contain zero *discord* [18–20], and are defined by Eq. (1) with the added constraints that the states $|\psi_i\rangle_A$ must be part of an orthonormal basis, and the same for states $|\phi_i\rangle_B$. Quantifying the amount of discord in a given state is also NP-complete [21], but it is harder than quantifying entanglement for small systems [22–24]. On the other hand, simply qualifying a state as discordant or not is a straightforward task [22,25]. The set of nondiscordant states is a subset of separable states. Currently, entanglement is the accepted, standard metric for quantum correlations, yet it should be noted that discord has one important advantage over entanglement, namely that it has a much stronger relationship to measurable quantities [26], while the detection of entanglement generally requires quantum tomography [27,28].

In the following, we would like to go beyond the typical scheme of describing the correlations in quantum states using analytical metrics and propose the study of quantum correlations using machine learning (ML) models forcing them to learn similarities between different classes of quantum states. We use neural networks (NNs) which have proven their usefulness in many applications, such as computer vision or signal processing [29,30], but recently also in quantum physics, e.g., in many-body theory [31–34], in entanglement modeling [35–40], or in state reconstruction [41]. However, in our approach, we carefully design the model architecture to be separability preserving and qubit permutations equivariant, which is part of the general trend of building NN models that respect symmetries of a given problem, for example, being rotation and translation group equivariant [42–45].

We present results where the ML model is first trained to distinguish between separable and entangled states (on both mixed and pure states) and then tested on similar sets. These sets (both in the training and testing phases), apart from entangled states, contain different variants of separable states, with special emphasis lain on discordant separable states. While the ML model trained only on pure states can only distinguish between product and nonproduct mixed states, hence signifying correlations, but not being able to discriminate between classical, quantum-classical (discord), and quantum-quantum (entanglement) correlations, quite surprisingly, we find that the performance of the separator trained also on mixed states is much better at signifying quantum discord than entanglement. This corroborates the intuition that a type of quantum correlations which are defined on the properties of measurement outcomes, as opposed to correlations defined on specifics of state preparation, should be more evident in a given state, even though the ML model was not trained directly on nondiscordant states (but on separable ones).

II. NEURAL NETWORK MODEL

Since the goal of using the neural network is not only to classify the discordance or entanglement of states, but also to produce a similarity measure, we propose to train the model in the unsupervised anomaly detector scheme [46]. Anomaly detection is a standard problem in ML where a model is trained to reconstruct normal data and then fails when trying to reconstruct abnormal data (anomaly). The reconstruction error, i.e., the model “loss function,” measures the similarity of a given unknown sample to the training data, and if it is greater than some assumed threshold, then the sample is treated as an anomaly. In our case, anomalies are states that are quantum correlated. The model is trained to recognize separable states of three qubits. We study three-qubit states to have a nontrivial yet manageable Hilbert space, keeping in mind that the proposed model construction can be easily extended to larger qubit registers.

To build the model, we use convolutional NNs (with a discrete convolution operation [47–49]) as they are commonly applied to build robust representations of correlations on data on regular grids such as matrices. First, we need to organize the convolution layers, so that the resulting NN can separate the input density matrix into single-qubit matrices. The constructed NN should preserve pure separable (product) states by definition and also not change the separability of mixed states. A closer look at the model scheme presented in Fig. 1 shows that the convolution layers are organized in a way that is a generalization of the partial trace operation, in the sense that the kernel simplified to the identity matrix would calculate the partial trace over the remaining qubits. Such an architecture implies qubit permutation equivariance, meaning that permutation of qubits in the input matrix gives the same permutation of the reconstructed state and preserves separability of any reconstructed state.

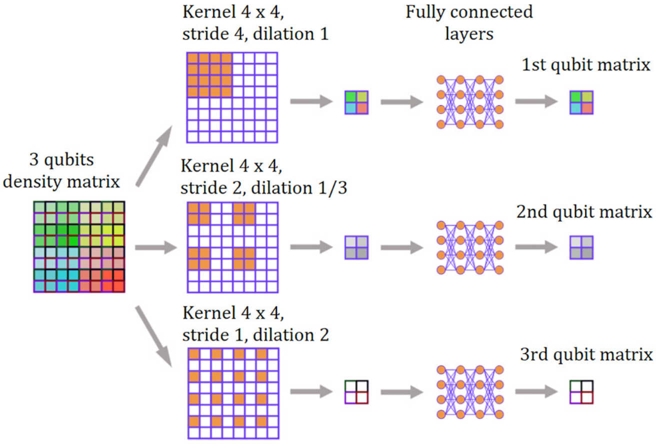

FIG. 1. The three-qubit neural network *separator* model. The input density matrix is convolved separately with three convolution layers, each with 4×4 kernels that work in parallel. The *stride* and *dilation* parameters defining these convolution layers are arranged in such a way that they operate on each qubit subspace separately (details can be found in the Appendix A). After the convolution layers, shape-preserving fully connected layers are applied independently for each qubit matrix.

Additionally, to extend the model complexity we include four fully connected (FC) layers with nonlinear activation functions (multilayer perceptron) after the convolutional ones. This extension has a physical justification that will become clear later. Importantly, to preserve the separability property, these layers are applied independently to each block that represents the qubit matrix. Moreover, to allow the neural network to generalize and reconstruct also separable mixed states, more sets of convolutions, i.e., *channels*, have been used, giving a separable density matrix that is a sum over N_K channels. Details can be found in the Appendix A. The reconstructed density matrix is given by

$$\hat{\rho}_{ABC} = \frac{1}{N}\sum_{i=1}^{N_K} \hat{\rho}_{Ai} \otimes \hat{\rho}_{Bi} \otimes \hat{\rho}_{Ci}, \tag{2}$$

where we calculate the sum of the Kronecker product of output single-qubit matrices ($\hat{\rho}_{Ai}$, $\hat{\rho}_{Bi}$, $\hat{\rho}_{Ci}$) over different N_k channels, numbered by i. Note that the reconstructed matrix is separable, but, in general, it does not fulfill the additional criteria which would make it a nondiscordant state from definition. Since a perfect decoupling into single-qubit matrices is possible only for separable states, the minimal distance between the original and reconstructed density matrices, obtained by the training of the neural network, can be seen as a measure of the separability of states.

Both the original and reconstructed density matrices have to be included in the NN loss function, which describes how different the reconstructed state is from the original state. In quantum information theory, one of the natural choices for a difference between density matrices is the Bures distance [50], but numerical experiments showed much better efficiency for training our NN model with the well-established L^1 norm:

$$\mathcal{L}(\rho_{ABC}) = \frac{1}{d^2}\sum_{i,j}\left|\rho_{ABC}^{ij} - \hat{\rho}_{ABC}^{ij}\right|. \tag{3}$$

Here ρ_{ABC}^{ij} and $\hat{\rho}_{ABC}^{ij}$ denote the elements of the original and reconstructed density matrices labeled by the indices i and j, and d^2 is the number of matrix elements. Using this loss function, we train the NN model so that the total reconstruction loss, summed over all training examples, is minimized. Note that to train the network, which we will call a *separator*, we do not need the labeled data. Instead, we compare inputs with outputs, which means that our training scheme is fully unsupervised. These types of ML models are also called *autoencoders* (we comment on this in more detail in the Appendix A).

Importantly, by choice of the model architecture, the separator network is suited to reconstruct separable states; hence, for these states we expect the value of the loss function to be close to zero. In contrast, for entangled states, the model is not able to reconstruct the input, and the difference between the two matrices will result in a significantly higher value of the loss $\mathcal{L}$. In order to distinguish between separable and entangled states, we set the anomaly threshold τ in such a way that if $\mathcal{L}(\rho_{ABC})$ is greater than τ for a given state ρ_{ABC}, then this state is classified as entangled. Otherwise, it is marked as separable.

III. DATA GENERATION

The most critical aspect of training an ML model is having a well-balanced and diversified data set. The modern approach in ML is to synthesize data with already known properties (e.g., forming separate classes in case of classifier training). To achieve this goal, we select different methods for generating random density matrices. The main idea is to generate random states using the quantum circuits approach [38], but in order to have a well-diversified data set, we also use a technique based on sampling from the uniform Haar measure [51] and an additional parametrized method designed by us (generation details can be found in the Appendix E).

We construct a training set composed of 530 000 pure and mixed separable states including a class of nondiscordant states and a validation set of 50 000 is constructed in the same way. Furthermore, we generate two test sets involving both separable and entangled states. The first of those (S_{pure}) is constructed from 15 000 random pure separable states and 15 000 random pure entangled states. This data set is generated evenly using both the quantum circuit approach and sampling from the Haar measure. The second test set (S_{mixed}) includes 65 000 mixed states, constructed from pure states either by mixing different states with random probabilities or by reducing a larger multiqubit space to the desired few-qubit size via the partial trace.

To classify states as separable during the generation process, we use the *negativity* [52,53], an entanglement measure based on the PPT (positive partial transpose) criterion of entanglement [54,55], which is one of the few measures that can be found directly from the density matrix for systems larger than two qubits. It is defined as the absolute sum of the negative eigenvalues of a bipartite density matrix after partial transposition is performed on either of the subsystems. The generation methods for mixed entangled states are specifically adjusted not to produce bound entangled states [56], which cannot be detected by negativity. To qualify discord, we use the criterion of Ref. [25], which is very straightforward to check, especially if one of the subsystems is small (relevant details are given in the Appendix F). Incidentally, it is common that a state is discordant only with respect to one part of a bipartition, so a state has to contain no discord with respect to both subsystems to be classified as a nondiscordant state [20]. For both entanglement and discord, the lack of correlations is checked three times: between qubit i and the rest of the three-qubit system, with $i = A, B, C$.

TABLE I. Average separator loss for different subsets in pure S_{pure} and mixed S_{pure} test sets quantifying similarity between different types of states and separable ones.

	$\langle\mathcal{L}\rangle$ for different subsets			
Test set	Separable	Nondiscordant	Discordant	Entangled
S_{pure}	5.9×10^{-4}	5.9×10^{-4}	0.042	0.042
S_{mixed}	5.2×10^{-3}	1.0×10^{-3}	0.017	0.021

IV. MODEL TRAINING AND RESULTS

In ML, a single pass through the entire training set, during which we update model parameters (typically using the gradient of a loss function in the parameter space), is called an epoch. The separator model was trained for 20 epochs, in a way that the model parameters for the epoch with the lowest averaged reconstruction loss $\langle\mathcal{L}\rangle$ on the validation set were saved and used for evaluation (testing) purposes.

For the test set that contains only pure states, S_{pure}, the average loss $\langle\mathcal{L}\rangle$ for separable states is equal to 5.9×10^{-4}, while in the case of entangled states, it is much larger, giving $\langle\mathcal{L}\rangle = 0.042$. Obviously, for the discord the numbers are the same—see Table I. Choosing the threshold value τ to obtain the highest possible accuracy [balanced accuracy (BA), defined in the Appendix B] of the model predictions yields $\tau = 1.25\times10^{-3}$ with the corresponding classification accuracy of 99.63% for the whole S_{pure} set.

For the test set with mixed states, S_{mixed}, in the case of nondiscordant states, the average loss $\langle\mathcal{L}\rangle$ is only about 1.5 times higher than for uncorrelated pure states, meaning that the original and reconstructed states are the same to an equivalent extent, while for mixed separable states it is 40 times higher, signaling that, on average, mixed separable states differ significantly more from the reconstruction. The same behavior, with mixed discordant states more similar to pure separable than mixed separable, can also be observed in supervised ML models—further details can be found in the Appendix C. Moreover, discordant and entangled states display roughly half $\langle\mathcal{L}\rangle$ for mixed states compared to pure correlated states, which is in agreement with the fact that mixed states are on average less correlated than pure states. Overall, the difference between zero and nonzero discord

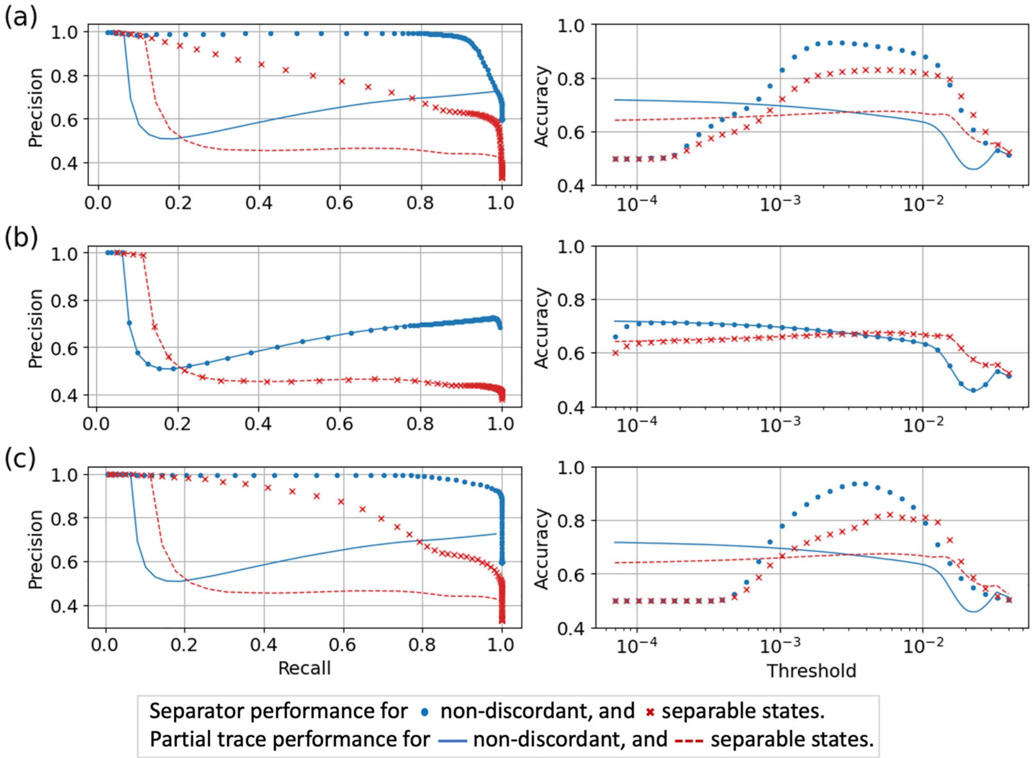

FIG. 2. (a) Separator performance for detecting discordant (blue dots) and entangled (red crosses) states, tested on S_{mixed}. Left: Precision vs recall curves. Right: Balanced accuracy, depending on the threshold τ value. For comparison, partial trace-based baseline model performance is also presented by blue solid and red dashed curves, respectively. (b) Same as (a) but for the separator model with removed FC layers—cf. Fig. 1. (c) Same as in (a) but for the separator trained on the nonproduct states only.

states (measured by reconstruction loss $\mathcal{L}$) is over an order of magnitude larger than between separable and entangled states.

Let us now test the separator performance using the mixed test set, S_{mixed}, and compare its practical usefulness for either discord or entanglement detection. We performed two independent tests. In the first, we took discord labels and tested the ability to classify discord, and in the second, we took separability labels for the same set S_{mixed} and tested the ability to classify entanglement. Since for mixed states a discordant state can also be separable, one cannot simultaneously balance nondiscordant vs discordant and separable vs entangled states in a single mixed test set; therefore, we use BA (see the Appendix B) as the model performance measure.

Figure 2 contains results of separator performance for detecting discordant (green curves) and entangled (red) states on the S_{mixed} set. Figure 2(a), right, shows that the model has much higher BA when detecting discord than entanglement in a wide range of thresholds, reaching 93% for $\tau = 0.0023$ in case of discord detection, while for entanglement a maximum of 82% is reached at $\tau = 0.0051$. Moreover, when training the model on the subspace of the original training domain where the product states have been excluded, as presented in Fig. 2(c), we clearly get better performance for discord in terms of the larger area under the precision-recall curve (left). Performance comparisons for training separately on some other classes of states can be found in the Appendix A.

The convolution part in the separator performs an operation that is a generalization of the partial trace. When removing the FC layers that extend the convolution layers, we observe that the model simply collapses to the partial trace baseline (example kernels for both cases can be found in the Appendices) and is analogous to what is obtained when training only on pure separable states. Figure 2(b) contains results for a separator with the FC layers removed: comparing accuracy with a partial trace baseline (blue and orange curves) we get practically the same accuracies.

We present the confusion matrices for the separator serving as a separability vs. entanglement classifier in Table II and a nondiscordance vs. discordance classifier in Table III. Overall, the separator tends to grossly overestimate the set of entangled states as seen in Table II, where for both thresholds presented ($\tau = 0.0023$ corresponding to the best discord classification and $\tau = 0.0051$ corresponding to the best entanglement classification) the number of separable states wrongly labeled as entangled is of the same order of magnitude as the number of correctly detected entangled states. The number of erroneously classified separable states is comparatively small. In

TABLE II. Confusion matrix for separable vs. entangled states classification at $\tau = 0.0023$ and 0.0051.

Predictions Labels		Separable	Entangled
Separable	$\tau = 0.0023$	28340	15257
Entangled	$\tau = 0.0023$	305	21098
Separable	$\tau = 0.0051$	30696	12901
Entangled	$\tau = 0.0051$	942	20461

TABLE III. Confusion matrix for nondiscordant vs. discordant states classification at $\tau = 0.0023$ and 0.0051.

Predictions Labels		Nondiscordant	Discordant
Nondiscordant	$\tau = 0.0023$	25114	1192
Discordant	$\tau = 0.0023$	3531	35163
Nondiscordant	$\tau = 0.0051$	25822	484
Discordant	$\tau = 0.0051$	5816	32878

contrast, the set of discordant states tends to be underestimated, with a discordant state being more likely to be labeled as not correlated than the other way around. This occurs to a much lesser extent than the overestimation of entanglement as seen in Table III. The results here are more balanced, with erroneous classifications averaging at a number an order of magnitude lower than the correct predictions.

V. CORRELATION MAP

In order to acquire a deeper understanding of how well the sets of nondiscordant and separable states are reconstructed by the ML classifier at different thresholds, we analyze a family of three-qubit states which are easy to parametrize, and study a two-dimensional (2D) diagram forming a map with regions differing in the type of correlations present (or absent) as seen in the inset of Fig. 3.

The family of states is constructed as follows. We take a separable mixed state of the form $\rho = p\rho_1 + (1-p)\rho_2$, where probability p is one of the parameters, and the density matrices ρ_1 and ρ_2 correspond to three-qubit pure states, $|\Psi_i\rangle = \bigotimes_k |\psi_i\rangle$, with $k = A, B, C$ and

$$|\psi_1\rangle = a|0\rangle + \sqrt{1-a^2}|1\rangle, \tag{4a}$$

$$|\psi_2\rangle = e^{-i\frac{\phi}{2}}\sqrt{1-a^2}|0\rangle - ae^{i\frac{\phi}{2}}|1\rangle, \tag{4b}$$

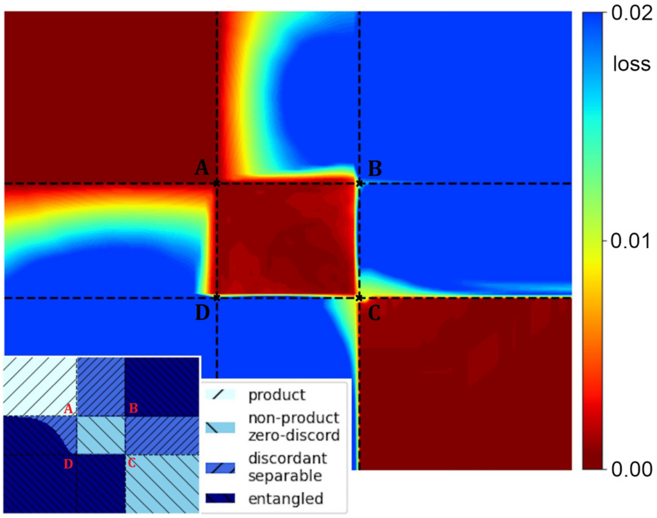

FIG. 3. Reconstruction loss $\mathcal{L}$ for the already trained separator model tested on the family of three-qubit states with their separability and discordance conditions known and parametrized on a 2D map (see inset for the map division into the classes and the Appendix D for detailed parametrization of the classes).

respectively. Coefficient a and phase factor ϕ constitute two other parameters that characterize the family of states. It now contains only separable states. For $p = 0, 1$ the state is pure (and consequently of product form), while it is mixed nondiscordant for $\phi = 0$ or $a = 0, 1$. To introduce entanglement into the system we find the spectral decomposition of $\rho = \sum_i \lambda_i|\varphi_i\rangle\langle\varphi_i|$ and increase the largest eigenvalue, λ_0:

$$\rho \rightarrow \frac{\sum_{i\neq 0}\lambda_i|\varphi_i\rangle\langle\varphi_i| + (\lambda_0 + c)|\varphi_0\rangle\langle\varphi_0|}{\sum_i \lambda_i + c}, \tag{5}$$

making c the final parameter which characterizes this family of states.

This allows us to plot the loss function (3) for selected cross sections of the Hilbert space and visualize how surfaces of equal $\mathcal{L}$ compare to the sets of nondiscordant and separable states. In Fig. 3 only the darkest regions in the inset are entangled; all other regions are separable and are further divided into product states, nondiscordant mixed states, and discordant separable states. The respective regions are obtained by careful choice of parameters p, a, ϕ, and c along the horizontal and vertical axes.

The main plot in Fig. 3 shows the loss $\mathcal{L}$ on this parameter space. The specifics of how the different parameters change in each region are described in detail in the Appendix D. It is noticeable that, while the partial trace-based model is capable only of distinguishing between product and nonproduct states (we show the plot and discuss this in detail in the Appendix A), the neural network distinguishes all nondiscordant states. The results presented in Fig. 3 are in agreement with the general results of Fig. 2, and Tables II and III, demonstrating why the discord detection results in false negatives (overestimating the set of nondiscordant states), while entanglement detection results in so many false positives.

VI. CONCLUSIONS

In conclusion, we designed and trained a neural network to distinguish between separable and entangled states on a set of mixed and pure states. For optimum threshold values, the network performs with 82% balanced accuracy for mixed states, as compared to over 99% classification accuracy of the same network for pure state entanglement. To understand this discrepancy outside of the trivial acknowledgment that determining mixed state entanglement is a much more complicated task than pure state entanglement, we have compared the results obtained on the mixed test set with other classes of uncorrelated states. We have found that the set of states labeled as uncorrelated by the separator overlaps best with the set of nondiscordant states. This is confirmed by the 93% balanced accuracy of detecting such states, outperforming the detection of both separable and product states. This result is surprising, because it shows that, for two different ML models (supervised and unsupervised), random mixed nondiscordant states are much more similar to states from a random set of pure separable states (as the lowest loss is observed for such states) than random mixed separable states. We supplement plots of ML metrics and confusion matrices with a map of numerically detected quantum correlations for cross sections of the three-qubit Hilbert space, which allows us to understand the nature of the sets of states which are qualified

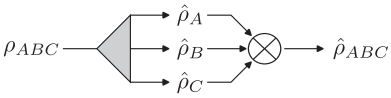

FIG. 4. The separator autoencoder composed of the trainable encoder neural network (gray triangle) and the analytical decoder that simply calculates the Kronecker product.

as uncorrelated by ML. The surface where the loss is near zero (no quantum correlations) reproduces the set of nondiscordant states extremely well, while the discordant separable states are largely classified as quantum correlated, illustrating why the separator grossly overestimates entanglement while only marginally underestimating discordance. The network architecture is designed carefully to preserve separability (i.e., it does not introduce any entanglement when reconstructing the output) and its output is equivariant with respect to qubit permutations. We observe that the simplified version of the separator model, which realizes just generalization of a partial trace operation, simply collapses to partial trace during the training. However, when extending it to a neural model capable to learn other representations by adding fully connected layers working separately within each qubit subspace, it learns more complex patterns and improves its performance over a partial trace baseline.

ACKNOWLEDGMENTS

M.K. and J.P. acknowledge support from the National Science Centre (Poland) under Grant No. 2021/43/D/ST3/01989. M.M.M. acknowledges that this research is part of Project No. 2021/43/P/ST3/03293 cofunded by the National Science Centre and the European Union's Horizon 2020 research and innovation program under Marie Skłodowska-Curie Grant No. 945339.

APPENDIX A: DETAILS ON THE SEPARATOR NEURAL NETWORK

In the following section, we analyze the structure of the NN model trained to reconstruct separable states and detect nonseparable states as an anomaly. Anomaly detection is a standard problem in ML where a model is trained to reconstruct normal data and then fails when trying to reconstruct abnormal data (anomaly). The reconstruction error, i.e., the model loss function, measures the similarity of a given unknown sample to the training data, and if it is greater than some assumed threshold, then the sample is treated as an anomaly. It is easy to construct a many-qubit density matrix as a product of single-qubit density matrices. However, the opposite operation is much more complicated. Then, this task can be seen as an *inverse problem*. Here, we propose to solve it with the help of a modified autoencoder-type NN, schematically depicted in Fig. 4. The typical autoencoder is composed of two connected networks: the *encoder* and the *decoder*. Usually, the output of the encoder is narrower than its input, so this part compresses the data. Next, the decoder decompresses it, and both parts of the autoencoder are trained to have the output as close as possible to the input. Our idea is to use the encoder to transform three-qubit density matrices into three single-qubit matrices in such a way that their product, calculated by the decoder, is as close to the original density matrix as possible. The autoencoder will be implemented as a NN, with only the encoder part trainable. The decoder simply calculates the Kronecker product of the single-qubit density matrices, as described in Eq. (2).

The backbone of the encoder is composed of convolution layers, which map the input density matrix into single-qubit subspaces. They are followed by FC layers that extend the model and introduce nonlinearity; however, they do not introduce entanglement, since they are applied independently for each of the reconstructed qubits.

Let us focus on a $d = 3$ qubit system ρ_{ABC}, bearing in mind that the proposed reconstruction can be easily extended to larger qubit registers. We require a neural network model $\mathcal{N}$, which we call a separator, to have the following properties:

$$\mathcal{N}(\rho_A \otimes \rho_B \otimes \rho_C) = c\,\rho_A \otimes \rho_B \otimes \rho_C, \tag{A1}$$

$$\mathcal{N}\left(\sum_i p_i\,\rho_{Ai} \otimes \rho_{Bi} \otimes \rho_{Ci}\right) = \sum_i \hat{p}_i\,\hat{\rho}_{Ai} \otimes \hat{\rho}_{Bi} \otimes \hat{\rho}_{Ci}, \tag{A2}$$

with some multiplicative constant c. To do so one can adjust the convolution layers $\mathcal{C}^{s,l}$, with kernel $\mathcal{K}$, step size (stride) s, and dilation l:

$$\mathcal{C}_{s,l}(\rho)^{i,j} = (\mathcal{K} *_l \rho)^{si,sj} = \sum_{k,q} \mathcal{K}^{k,q} \rho^{si+l(k),sj+l(q)}, \tag{A3}$$

where $i, j = 0, 1$, and $k, q = 0, \ldots, 2^{d-1} - 1$, so that the kernels overlap with the respective blocks of the density matrix and therefore convolving them against the density matrix results with output matrices representing systems with a reduced number of qubits. Dilation can be just a simple scaling of the sum index: $l(k) = l\,k$, with rate l meaning that the kernel only touches the signal at every lth entry, or more complex indexing.

Specifically, choosing $\mathcal{C}_A = \mathcal{C}_{4,1}$ with kernel $\mathcal{K}$ of size 4×4 allows the neural network to learn the structure of repeating blocks $\rho_B \otimes \rho_C$ and therefore to reduce the input density matrix to

$$\mathcal{C}_{4,1}(\rho_A \otimes \rho_B \otimes \rho_C) = c\,\rho_A,$$

with c being some multiplicative constant, as in Eq. (A1). To extract the last one, we need $\mathcal{C}_C = \mathcal{C}_{1,2}$. To extract the second qubit, more complex dilation is needed in $\mathcal{C}_B$. All kernels are visualized in Fig. 1. The input density matrix, divided into two channels (respectively for the real and imaginary parts), is convolved separately with three convolution layers, each with 4×4 kernels. The first one is set up using a stride equal to 4, as a result producing a matrix corresponding to the first qubit subspace. In the second convolution, the kernel is divided into four blocks of size 2×2, separated with dilations resulting in the second qubit subspace. The third convolution consists of the kernel dilated with extra space between every weight, therefore extracting the information about the last qubit. Finally, after applying all the convolutions, one can combine them and construct the output separable density matrix $\mathcal{C}_A(\rho_{ABC}) \otimes \mathcal{C}_B(\rho_{ABC}) \otimes \mathcal{C}_C(\rho_{ABC}) = \hat{\rho}_A \otimes \hat{\rho}_B \otimes \hat{\rho}_C$.

It is obvious that in this way the defined separator model $\mathcal{N}$ fulfils Eq. (A1). However, to enable the neural network to generalize and reconstruct separable mixed states, i.e., obey Eq. (A2), a larger number of (triples of) kernels is chosen (we use $N_K = 24$ kernels in each type of convolution layer). To simplify the implementation and keep neural network values real, we doubled the number of kernels, that is, $2N_K$, so half of the channels processed the real and the other half processed the imaginary part of the density matrix. At this point, we also add four FC layers for each qubit (A, B, C) separately to extend the complexity of the model. Subsequently, we calculate the sum of the Kronecker product of output single-qubit matrices $(\hat{\rho}_{Ai}, \hat{\rho}_{Bi}, \hat{\rho}_{Ci})$ over different kernels numbered by i and arrive at the reconstructed density matrix as in Eq. (2).

Both the original and reconstructed density matrices are included in the loss function of the neural network defined in Eq. (3). Using this loss function, we train the separator model so that the total reconstruction loss, summed over all training examples, is minimized. Note that to train the network, we do not need the data to be labeled. Instead, we simply compare inputs with outputs, and therefore, our model scheme is fully unsupervised. The trainable weights of the kernels are adjusted using the procedure called gradient descent, where the gradients of the loss are calculated with respect to the weights. Subsequently, all parameters of the model, e.g., each kernel weight $\mathcal{K}_{ij}$, are updated as follows:

$$\mathcal{K}^{ij}_{\text{new}} = \mathcal{K}^{ij}_{\text{old}} - \alpha \frac{\partial \mathcal{L}}{\partial \mathcal{K}^{ij}_{\text{old}}}, \tag{A4}$$

where α is the learning rate. We proceed similarly for weights in FC layers. This procedure is repeated until the reconstruction loss [Eq. (3)] reaches (we hope) the global minimum, which means that the model is already trained. Importantly, by the choice of the model architecture, the separator network is suited to reconstruct separable states, and for these states, we expect the value of the loss function to be close to zero. In contrast, in the case of an entangled state, the model will not be able to reconstruct the input, and the output density matrix should differ from the input one, resulting in a significantly higher value of the loss. Therefore, to distinguish between separable and entangled states, we set the anomaly threshold τ in such a way that if $\mathcal{L}(\rho_{ABC})$ is larger than τ for a given state ρ_{ABC}, then this state is classified as entangled and, otherwise, it is marked as separable. Interestingly, in addition to classification purposes, the loss value $\mathcal{L}(\rho_{ABC})$ can also be interpreted as a similarity measure between the states used in the training dataset and given tested state ρ_{ABC}. In Fig. 5 we present the extended results for the separator model with and without four FC layers trained on different subsets of the original training set. As before, the neural network is evaluated on the S_{mixed} set. We start to train the model in the simplest possible scenario where the dataset consists of 190 000 pure separable states. The model trained on such a dataset collapses to the baseline trace model. It is expected behavior since in this case the baseline model is a sufficient criterion for distinguishing between entangled and separable states. The same situation occurs when we extend our training set to the wider set of product states by adding 120 000 mixed product states. Interestingly, in this case, the neural network reflects partial trace operation even more precisely, which is revealed by a wider range of thresholds producing overlapping accuracy. Including FC layers, and thus introducing the nonlinearity, slightly limits this phenomenon; however, since the baseline model is still a precise criterion of separability in the case of product states, hence the neural network is not penalized for not learning more complex operations. The situation changes when we append 100 000 nonproduct nondiscordant states. Now, the baseline is no longer reliable and the network needs to learn extra features in order to correctly separate and reconstruct such states. Notably, the separator model is capable of achieving this task only when extra FC layers are present. Otherwise, it collapses to the baseline. This situation reoccurs for the full training

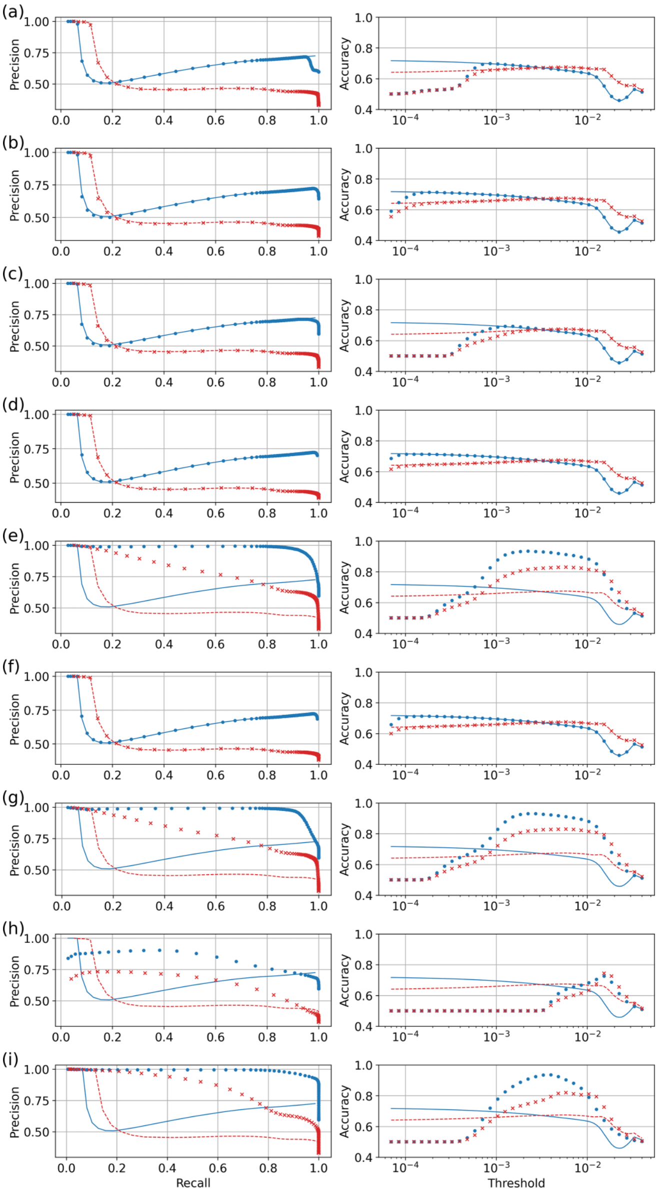

FIG. 5. Same as in Fig. 2, but for the separator trained only on some part of the original training set: (a) pure separable states, (b), (c) product states, (d), (e) nondiscordant states, (f), (g) separable states, and (h), (i) nonproduct separable states. Results are shown for the (standard) separator with four fully connected layers after convolutional ones (c), (e), (g), (i) and without it, containing only convolutional layers (a), (b), (d), (f), (h).

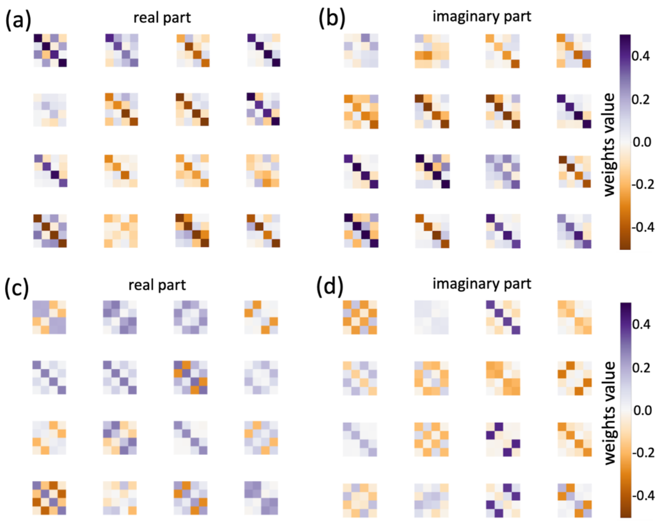

FIG. 6. Example weights of 4×4 convolutional kernels for the already trained three-qubit separator model. The kernels reconstruct (a) real and (b) imaginary parts of the first qubit (weights for two other qubits are similar) for the plain separator model, with just convolutional layers. Similarly, (c) and (d) show real and imaginary parts of kernels for the separator with additional fully connected layers.

set consisting of 530 000 separable states. For the sake of completeness, we additionally test our model after training it on the reduced domain of nonproduct separable states. The data set in this case is constructed separately from 360 000 states, which are generated with the same methodology as for the original training set. The main difference of the model trained on such a subdomain is that even without any FC layers it does not collapse to a partial trace operation, which may suggest that using only nonproduct separable states for training can also be beneficial for the model with FC layers. In Fig. 6 we present exemplary kernels learned for the separator model with (right) and without FC layers (left) trained on the full data set (as in the main text). The model with only convolutional layers collapses to a partial trace baseline which manifests itself in the fact that its kernels resemble identity matrices (the partial trace is equivalent to the convolutional kernel being equal to the identity matrix).

In Fig. 7 we also present a reconstruction loss map analogous to the map in Fig. 3 in the main text, but obtained using a partial trace as a baseline replacement for the separator model. On this map, only product states are clearly detected as not containing correlations, as is to be expected; however, the region in the lower right square of the map, which represents some space of mixed zero-discord states, is signified as containing fewer correlations. Note that the other region of similar type (in the center of the plot) is not correspondingly distinguished.

APPENDIX B: CLASSIFICATION METRICS

In machine learning to evaluate classification or detection results, we typically use various types of metrics that measure model performance.

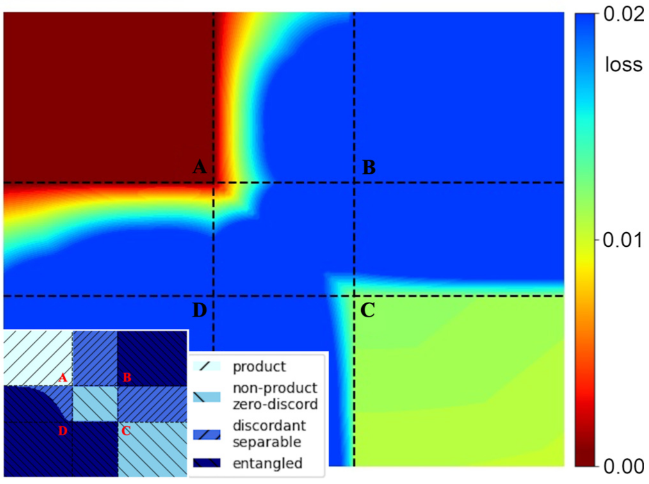

FIG. 7. Reconstruction loss $\mathcal{L}$ for the partial trace-based model tested on a family of three-qubit states with their separability and discordance conditions parametrized on a 2D map (see inset for the map division into the classes and Appendix D for detailed parametrization of the classes).

The standard metrics [57] used in binary classification, also to tune the model hyperparameters, e.g., classification threshold τ, are precision $\mathrm{PR} = \frac{\mathrm{TP}}{\mathrm{TP+FP}}$ and recall $\mathrm{RC} = \frac{\mathrm{TP}}{\mathrm{TP+FN}}$. Here TP, TN, FP, and FN denote respectively the number of true and false positives and negatives; by “negatives” we mean nondiscordant (or separable), while by “positives” we mean discordant (or entangled) states.

The procedure called “precision-recall tradeoff” involves choosing a threshold that maximizes the area under the precision-recall curve. However, due to the fact that our dataset is imbalanced (we cannot have balanced separable or entangled and nondiscordant or discordant classes at the same time), we use a metric typically defined for imbalanced data sets. Balanced accuracy is defined as the average of recall obtained on each class [58]: $\mathrm{BA} = \frac{1}{2}(\frac{\mathrm{TP}}{\mathrm{TP+FN}} + \frac{\mathrm{TN}}{\mathrm{TN+FP}})$.

APPENDIX C: SIMILARITY BETWEEN STATES

In order to verify whether the statement from the main text, that the mixed nondiscordant states are much more similar to states from a set of pure separable states (as the lowest loss is observed for such states in Table I) than mixed separable states, also holds for other machine learning models, we tested two different models, as introduced in Ref. [38]. Both of them, CNN and Siamese CNN, were trained in a supervised manner to classify if a given three-qubit state is entangled [38]. In Table IV there are presented average output probabilities for different classes of states in the test set for both models. We can clearly observe that the outputs show the same similarities between states as for the unsupervised model, used in the main text.

APPENDIX D: PARAMETER VARIATIONS ON A THREE-QUBIT MAP

The family of states in the map in Fig. 3 is constructed as follows. We take a separable mixed state of the form

TABLE IV. Average output probability p for CNN (top) and Siamese CNN (bottom) models trained to predict if a given state is entangled.

	$\langle p\rangle$ for CNN model			
Test set	Separable	Nondiscordant	Discordant	Entangled
S_{pure}	0.104	0.104	0.968	0.968
S_{mixed}	0.221	0.115	0.686	0.887
	$\langle p\rangle$ for Siamese CNN model			
Test set	Separable	Nondiscordant	Discordant	Entangled
S_{pure}	0.027	0.027	0.969	0.969
S_{mixed}	0.105	0.038	0.570	0.824

$\rho = p\rho_1 + (1-p)\rho_2$, where probability p is one of the parameters, and the density matrices ρ_1 and ρ_2 correspond to three-qubit pure states, $|\Psi_i\rangle = \bigotimes_k |\psi_i\rangle$, with $k = A, B, C$ and

$$|\psi_1\rangle = a|0\rangle + \sqrt{1-a^2}|1\rangle, \tag{D1a}$$

$$|\psi_2\rangle = e^{-i\frac{\phi}{2}}\sqrt{1-a^2}|0\rangle - ae^{i\frac{\phi}{2}}|1\rangle, \tag{D1b}$$

respectively. Coefficient a and phase factor ϕ constitute two other parameters that characterize the family of states. It now contains only separable states. For $p = 0, 1$ the state is pure (and consequently of product form), while it is mixed nondiscordant for $\phi = 0$ or $a = 0, 1$. To introduce entanglement into the system we find the spectral decomposition of $\rho = \sum_i \lambda_i|\varphi_i\rangle\langle\varphi_i|$ and increase the largest eigenvalue, λ_0:

$$\rho \to \frac{\sum_{i\neq 0}\lambda_i|\varphi_i\rangle\langle\varphi_i| + (\lambda_0 + c)|\varphi_0\rangle\langle\varphi_0|}{\sum_i \lambda_i + c}, \tag{D2}$$

making c the final parameter which characterizes this family of states.

Now we detail how the specific parameters are changed in the different regions of the map to generate states corresponding to different classes in terms of quantum and classical correlations. Specifically, we start from point A, where the parameter $a = 1/\sqrt{2}$ takes the maximal value, while all others are set to minimal values: $p = 0$, $\phi = 0$, and $c = 0$. Moving along the axis AB we increase the parameter p so that it takes the maximal value $1/2$ in point B. Similarly, coefficient a is decreased along the axis BC, so that in point C it is equal to zero. Moving outside the square region $ABCD$ we fix parameters a and p to the boundary values and start increasing the parameters ϕ and c. The phase ϕ grows linearly along both horizontal and vertical axes in such a way that inside the square $ABCD$ it is equal to zero and it is equal to $\pi/2$ at the edges of the map. The parameter c is increased along the counterdiagonal in the top right square, taking the minimal value zero in point B and maximal value 1 in the top right corner of the map. The parametrization is performed symmetrically on both sides of the diagonal of the map. Therefore, the axis AB in the upper right triangle corresponds to the axis AD in the lower left triangle. Similarly, BC is reflected to DC. Keeping that in mind, we note that parameter c is additionally increased by Δc in the lower left triangle. This extra term grows linearly along the counterdiagonal from $\Delta c = 0$ in the center of square $ABCD$ to $\Delta c = 1$ in the bottom left corner of the map.

APPENDIX E: PARAMETRIZED METHOD OF GENERATING THREE-QUBIT DENSITY MATRICES

The most critical aspect of training an ML model is having a well-balanced and diversified data set. The modern approach in ML is to synthesize data with already known properties (e.g., forming separate classes in case of classifier training). To achieve this goal, we select different methods for generating random density matrices. The main idea is to generate random states using the quantum circuits approach [38], but in order to have a well-diversified data set, we also use a technique based on sampling from the uniform Haar measure [51].

We construct a training set composed of 530 000 pure and mixed separable states including a class of nondiscordant states and a validation set of 50 000 is constructed in the same way. Furthermore, we generate two test sets involving both separable and entangled states. The first of those (S_{pure}) is constructed from 15 000 random pure separable states and 15 000 random pure entangled states. This data set is generated evenly using both the quantum circuit approach and sampling from the Haar measure. The second test set (S_{mixed}) includes 65 000 mixed states, constructed from pure states either by mixing different states with random probabilities or by reducing a larger multiqubit space to the desired few-qubit size via the partial trace.

Now we describe the original parametrized method used to generate random mixed states when building the training set. To generate such states, we also used the quantum circuits approach and a technique based on sampling from the uniform Haar measure—both are described in Ref. [38].

Let us start by defining a three-qubit pure separable state:

$$\begin{aligned}|\psi_{\text{sep}}\rangle &= |\psi_1\rangle \otimes |\psi_2\rangle \otimes |\psi_3\rangle\\ &= (a_1|0\rangle + b_1|1\rangle) \otimes (a_2|0\rangle + b_2|1\rangle) \otimes (a_3|0\rangle + b_3|1\rangle)\\ &= a_1a_2a_3|000\rangle + a_1a_2b_3|001\rangle + a_1a_3b_2|010\rangle\\ &\quad + a_1b_2b_3|011\rangle + a_2a_3b_1|100\rangle + a_2b_1b_3|101\rangle\\ &\quad + a_3b_1b_2|110\rangle + b_1b_2b_3|111\rangle,\end{aligned} \tag{E1}$$

where the complex coefficients of each state fulfill

$$|a_i|^2 + |b_i|^2 = 1. \tag{E2}$$

We then introduce entanglement into the system for each pair of qubits via phases ϕ_{ij}:

$$\begin{aligned}|\psi_{\text{ent}}\rangle &= a_1a_2a_3|000\rangle + a_1a_2b_3|001\rangle + a_1a_3b_2|010\rangle\\ &\quad + a_1b_2b_3e^{i\phi_{23}}|011\rangle + a_2a_3b_1|100\rangle\\ &\quad + a_2b_1b_3e^{i\phi_{13}}|101\rangle + a_3b_1b_2e^{i\phi_{12}}|110\rangle\\ &\quad + b_1b_2b_3e^{i(\phi_{12}+\phi_{13}+\phi_{23})}|111\rangle.\end{aligned} \tag{E3}$$

Finally, the decoherence coefficients c_i are included for each qubit separately in order to obtain mixed states, mimicking the operation of single-qubit dephasing channels:

$$\rho_i = |\psi_i\rangle\langle\psi_i| = \begin{pmatrix}|a_i|^2 & a_ib_i^*\\ a_i^*b_i & |b_i|^2\end{pmatrix} \tag{E4}$$

$$\xrightarrow{\text{decoherence}} \begin{pmatrix}|a_i|^2 & c_ia_ib_i^*\\ c_i^*a_i^*b_i & |b_i|^2\end{pmatrix} = \tilde{\rho}_i. \tag{E5}$$

Combining all of the above, one arrives at the following:

$$\rho = |\psi_{\mathrm{ent}}\rangle\langle\psi_{\mathrm{ent}}| \xrightarrow{\text{decoherence}} \tilde{\rho}, \tag{E6}$$

where the density matrix $\tilde{\rho}$ is parametrized with independent coefficients a_i and c_i as well as phases ϕ_{ij}. In order to obtain a diversified set of states, we propose to choose these parameters uniformly at random over the interval [0, 1] for a_i and c_i, and over the interval $[0, 2\pi]$ for phases ϕ_i. Moreover, we randomize the states by evolving them with local single-qubit gates:

$$U(\theta, \phi, \lambda) = \begin{pmatrix} \cos\frac{\theta}{2} & -e^{i\lambda}\sin\frac{\theta}{2} \\ e^{i\phi}\sin\frac{\theta}{2} & e^{i(\phi+\lambda)}\cos\frac{\theta}{2} \end{pmatrix}, \tag{E7}$$

where θ, ϕ, and λ are some arbitrary Euler angles. Hence, sampling the angles randomly over the interval $[0, 2\pi]$, independently for separate qubits, one can write the final randomized density matrix as

$$\tilde{\rho}_{\mathrm{rand}} = (U_1 \otimes U_2 \otimes U_3)\tilde{\rho}(U_1 \otimes U_2 \otimes U_3)^{\dagger}. \tag{E8}$$

APPENDIX F: CLASSIFYING STATES AS NONDISCORDANT

To classify a state as discordant or not, we use an if-and-only-if criterion for nondiscordant states from Ref. [25]. The criterion states that a bipartite state with subsystems of arbitrary dimensions N and M has zero quantum discord with respect to the system of dimension M, if and only if all blocks of the $(NM)\times(NM)$ density matrix, which are obtained by partitioning the density matrix into N^2 square matrices of dimension M, are normal matrices and commute with each other.

The required partition is performed starting from the density matrix written in the form

$$\hat{\sigma} = \sum_{kq}\sum_{nm} P_{kq}^{nm}|k\rangle\langle q| \otimes |n\rangle\langle m|, \tag{F1}$$

where the indices and corresponding states k and q correspond to the system of dimension N, while the indices and states n and m correspond to the subsystem of dimension M. The partition into blocks requires leaving only the elements which correspond to matrix elements $|k\rangle\langle q|$ in the subspace of the system of dimension N:

$$\hat{\sigma}_{kq} = \langle k|\hat{\sigma}|q\rangle. \tag{F2}$$

Each choice of k and q yields a different matrix.

A normal matrix is one that commutes with its Hermitian conjugate (the matrix obtained through the partition does not need to be a density matrix), thus fulfilling

$$[\hat{\sigma}_{kq}, \hat{\sigma}_{kq}^{\dagger}] = 0, \tag{F3}$$

whereas commutation obviously requires that for all k and q and for all k' and q' one has

$$[\hat{\sigma}_{kq}, \hat{\sigma}_{k'q'}] = 0. \tag{F4}$$

Both criteria are fulfilled, if and only if the state has zero discord with respect to the subsystem of dimension M.

It is relevant to keep in mind that a zero discord state must not have any correlations of this type with respect to the subsystems of both size M and size N, so in the case of our three-qubit system and each of the three-qubit–two-qubit partitions we will require two checks for zero discord, one with respect to the qubit and the other with respect to the two-qubit subsystem. In total, this means that checking for the presence of discord in a given three-qubit state is performed six times due to the different partitions and the asymmetry of the discord.

APPENDIX G: NUMERICAL METHODS

Data generation is performed with the usage of the QISKIT [59] library. All data sets can be generated in ≈8 h. The neural network model (encoder and decoder part) is implemented using the PYTORCH [60] library with an automatic differentiation engine (autograd) used to track gradients of the model parameters. The training of the single model on the GPU Nvidia RTX 3070 takes ≈1 day.

[1] N. Gisin, G. Ribordy, W. Tittel, and H. Zbinden, Quantum cryptography, Rev. Mod. Phys. **74**, 145 (2002).

[2] S. Pirandola, U. L. Andersen, L. Banchi, M. Berta, D. Bunandar, R. Colbeck, D. Englund, T. Gehring, C. Lupo, C. Ottaviani, J. L. Pereira, M. Razavi, J. S. Shaari, M. Tomamichel, V. C. Usenko, G. Vallone, P. Villoresi, and P. Wallden, Advances in quantum cryptography, Adv. Opt. Photon. **12**, 1012 (2020).

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

[4] S. Wehner, D. Elkouss, and R. Hanson, Quantum internet: A vision for the road ahead, Science **362**, eaam9288 (2018).

[5] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University, Cambridge, England, 2000).

[6] B. Demirel, W. Weng, C. Thalacker, M. Hoban, and S. Barz, Correlations for computation and computation for correlations, npj Quantum Inf. **7**, 29 (2021).

[7] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters, Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels, Phys. Rev. Lett. **70**, 1895 (1993).

[8] D. Bouwmeester, J.-W. Pan, K. Mattle, M. Eibl, H. Weinfurter, and A. Zeilinger, Experimental quantum teleportation, Nature (London) **390**, 575 (1997).

[9] D. Boschi, S. Branca, F. De Martini, L. Hardy, and S. Popescu, Experimental realization of teleporting an unknown pure quantum state via dual classical and Einstein-Podolsky-Rosen channels, Phys. Rev. Lett. **80**, 1121 (1998).

[10] K. Roszak and J. K. Korbicz, Purifying teleportation, Quantum **7**, 923 (2023).

[11] M. Horodecki, P. Horodecki, R. Horodecki, J. Oppenheim, A. Sen, U. Sen, and B. Synak-Radtke, Local versus nonlocal information in quantum-information theory: Formalism and phenomena, Phys. Rev. A **71**, 062307 (2005).

[12] K. Modi, A. Brodutch, H. Cable, T. Paterek, and V. Vedral, The classical-quantum boundary for correlations: Discord and related measures, Rev. Mod. Phys. **84**, 1655 (2012).

[13] K. Roszak and Ł. Cywiński, The relation between the quantum discord and quantum teleportation: The physical interpretation of the transition point between different quantum discord decay regimes, Europhys. Lett. **112**, 10002 (2015).

[14] R. F. Werner, Quantum states with Einstein-Podolsky-Rosen correlations admitting a hidden-variable model, Phys. Rev. A **40**, 4277 (1989).

[15] M. B. Plenio and S. Virmani, An introduction to entanglement measures, Quantum Inf. Comput. **7**, 1 (2007).

[16] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, Quantum entanglement, Rev. Mod. Phys. **81**, 865 (2009).

[17] L. Gurvits, Classical complexity and quantum entanglement, J. Comput. Syst. Sci. **69**, 448 (2004).

[18] H. Ollivier and W. H. Zurek, Quantum discord: A measure of the quantumness of correlations, Phys. Rev. Lett. **88**, 017901 (2001).

[19] L. Henderson and V. Vedral, Classical, quantum and total correlations, J. Phys. A **34**, 6899 (2001).

[20] K. Modi, A pedagogical overview of quantum discord, Open Syst. Inf. Dyn. **21**, 1440006 (2014).

[21] Y. Huang, Computing quantum discord is NP-complete, New J. Phys. **16**, 033027 (2014).

[22] B. Dakić, V. Vedral, and Č. Brukner, Necessary and sufficient condition for nonzero quantum discord, Phys. Rev. Lett. **105**, 190502 (2010).

[23] A. Miranowicz, P. Horodecki, R. W. Chhajlany, J. Tuziemski, and J. Sperling, Analytical progress on symmetric geometric discord: Measurement-based upper bounds, Phys. Rev. A **86**, 042123 (2012).

[24] T. Tufarelli, T. MacLean, D. Girolami, R. Vasile, and G. Adesso, The geometric approach to quantum correlations: Computability versus reliability, J. Phys. A **46**, 275308 (2013).

[25] J.-H. Huang, L. Wang, and S.-Y. Zhu, A new criterion for zero quantum discord, New J. Phys. **13**, 063045 (2011).

[26] K. Roszak and L. Cywiński, Equivalence of qubit-environment entanglement and discord generation via pure dephasing interactions and the resulting consequences, Phys. Rev. A **97**, 012306 (2018).

[27] D. F. V. James, P. G. Kwiat, W. J. Munro, and A. G. White, Measurement of qubits, Phys. Rev. A **64**, 052312 (2001).

[28] M. Christandl and R. Renner, Reliable quantum state tomography, Phys. Rev. Lett. **109**, 120403 (2012).

[29] J. Chai, H. Zeng, A. Li, and E. W. Ngai, Deep learning in computer vision: A critical review of emerging techniques and application scenarios, Mach. Learn. Applic. **6**, 100134 (2021).

[30] M. Alam, M. Samad, L. Vidyaratne, A. Glandon, and K. Iftekharuddin, Survey on deep neural networks in speech and vision systems, Neurocomputing **417**, 302 (2020).

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

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

[33] Y. Levine, O. Sharir, N. Cohen, and A. Shashua, Bridging many-body quantum physics and deep learning via tensor networks, in *Mathematical Aspects of Deep Learning*, edited by P. Grohs and G. Kutyniok (Cambridge University Press, New York, 2022), pp. 439–474.

[34] X. Cheng, S. Zhang, P. C. H. Nguyen, S. Azarfar, G.-W. Chern, and S. S. Baek, Convolutional neural networks for large-scale dynamical modeling of itinerant magnets, arXiv:2306.11833.

[35] Y. Levine, O. Sharir, N. Cohen, and A. Shashua, Quantum entanglement in deep learning architectures, Phys. Rev. Lett. **122**, 065301 (2019).

[36] J. Roik, K. Bartkiewicz, A. Černoch, and K. Lemr, Accuracy of entanglement detection via artificial neural networks and human-designed entanglement witnesses, Phys. Rev. Appl. **15**, 054006 (2021).

[37] Y. Chen, Y. Pan, G. Zhang, and S. Cheng, Detecting quantum entanglement with unsupervised learning, Quantum Sci. Technol. **7**, 015005 (2022).

[38] J. Pawłowski and M. Krawczyk, Identification of quantum entanglement with siamese convolutional neural networks and semi-supervised learning, arXiv:2210.07410.

[39] N. Asif, U. Khalid, A. Khan, T. Q. Duong, and H. Shin, Entanglement detection with artificial neural networks, Sci. Rep. **13**, 1562 (2023).

[40] J. Ureña, A. Sojo, J. Bermejo, and D. Manzano, Entanglement detection with classical deep neural networks, arXiv:2304.05946.

[41] A. M. Palmieri, G. Müller-Rigat, A. K. Srivastava, M. Lewenstein, G. Rajchel-Mieldzioć, and M. Płodzień, Enhancing quantum state tomography via resource-efficient attention-based neural networks, arXiv:2309.10616.

[42] T. Cohen and M. Welling, Group equivariant convolutional networks, in *Proceedings of the 33rd International Conference on Machine Learning*, Proc. Mach. Learn. Res. Vol. 48 (PMLR, 2016), pp. 2990–2999.

[43] E. J. Bekkers, M. W. Lafarge, M. Veta, K. A. J. Eppenhof, J. P. W. Pluim, and R. Duits, Roto-translation covariant convolutional networks for medical image analysis, in *Medical Image Computing and Computer Assisted Intervention—MICCAI 2018*, edited by A. F. Frangi, J. A. Schnabel, C. Davatzikos, C. Alberola-López, and G. Fichtinger (Springer, New York, 2018), pp. 440–448.

[44] T. Cohen, M. Weiler, B. Kicanaoglu, and M. Welling, Gauge equivariant convolutional networks and the icosahedral CNN, in *Proceedings of the 36th International Conference on Machine Learning* (PMLR, 2019), pp. 1321–1330.

[45] Y. Nagai and A. Tomiya, Self-learning Monte Carlo with equivariant transformer, arXiv:2306.11527.

[46] P. Malhotra, L. Vig, G. Shroff, and P. Agarwal, Long short term memory networks for anomaly detection in time series, in *Proceedings of the European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning* (ESANN, 2015).

[47] M. D. Zeiler and R. Fergus, Visualizing and understanding convolutional networks, in *European Conference on Computer Vision* (Springer, New York, 2014), pp. 818–833.

[48] F. Yu and V. Koltun, Multi-scale context aggregation by dilated convolutions, in *Proceedings of the 4th International Conference on Learning Representations* (ICLR, 2016).

[49] V. Dumoulin and F. Visin, A guide to convolution arithmetic for deep learning, arXiv:1603.07285.

[50] A. Laha, A. Aggarwal, and S. Kumar, Random density matrices: Analytical results for mean root fidelity and the mean-square Bures distance, Phys. Rev. A **104**, 022438 (2021).

[51] F. Mezzadri, How to generate random matrices from the classical compact groups, Notices AMS **54**, 592 (2007).

[52] G. Vidal and R. F. Werner, Computable measure of entanglement, Phys. Rev. A **65**, 032314 (2002).

[53] M. B. Plenio, Logarithmic negativity: A full entanglement monotone that is not convex, Phys. Rev. Lett. **95**, 090503 (2005).

[54] A. Peres, Separability criterion for density matrices, Phys. Rev. Lett. **77**, 1413 (1996).

[55] M. Horodecki, P. Horodecki, and R. Horodecki, Separability of mixed states: Necessary and sufficient conditions, Phys. Lett. A **223**, 1 (1996).

[56] P. Horodecki, M. Horodecki, and R. Horodecki, Bound entanglement can be activated, Phys. Rev. Lett. **82**, 1056 (1999).

[57] D. Powers, Evaluation: From precision, recall and f-measure to ROC, informedness, markedness and correlation, J. Mach. Learn. Tech. **2**, 37 (2011).

[58] K. H. Brodersen, C. S. Ong, K. E. Stephan, and J. M. Buhmann, The balanced accuracy and its posterior distribution, in *Proceedings of the 2010 20th International Conference on Pattern Recognition* (IEEE, New York, 2010), pp. 3121–3124.

[59] H. Abraham *et al.*, Qiskit: An open-source framework for quantum computing (2019).

[60] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga *et al.*, Pytorch: An imperative style, high-performance deep learning library, in *Advances in Neural Information Processing Systems*, Vol. 32 (2019).