

Entanglement Classification via Witness Operators generated by Support Vector Machine

Claudio Sanavio,^{1,2,*} Edoardo Tignone,³ and Elisa Ercolessi^{1,2}

¹*Dipartimento di Fisica e Astronomia dell'Università di Bologna, I-40127 Bologna, Italy*

²*INFN, Sezione di Bologna, I-40127 Bologna, Italy*

³*Leithà S.r.l. | Unipol Group, Bologna, Italy*

Although entanglement is a basic resource for reaching quantum advantage in many computation and information protocols, we lack a universal recipe for detecting it, with analytical results obtained for low dimensional systems and few special cases of higher dimensional systems. In this work, we use a machine learning algorithm, the support vector machine with polynomial kernel, to classify separable and entangled states. We apply it to two-qubit and three-qubit systems, and we show that, after training, the support vector machine is able to recognize if a random state is entangled with an accuracy up to 92% for the two-qubit system and up to 98% for the three-qubit system. We also describe why and in what regime the support vector machine algorithm is able to implement the evaluation of an entanglement witness operator applied to many copies of the state, and we describe how we can translate this procedure into a quantum circuit.

I. INTRODUCTION

Quantum entanglement [1] is one of the main features that distinguish between classical and quantum states and represents one of the basic ingredients for reaching quantum advantage in computation and information protocols [2–4]. An N -particle state is entangled if its density operator can not be written as tensor product or sum of tensor products of N single-particle density operators or, in other words, if the state of the system cannot be entirely described considering its N single components only. Despite the simplicity of such a definition, the problem of identifying, classifying and quantifying entanglement is mathematically extremely hard. Many analytical and numerical results have been obtained in the study of bipartite entanglement, including the well known positive partial transpose (PPT) criterium, which establishes that, for two-qubit and qubit-qutrit states, the positivity of the partial transposition of the density operator provides a necessary [5] and sufficient [6] condition for entanglement. In spaces of higher dimension, however, the PPT criterion offers only a necessary condition for separability [5], reflecting the fact that entanglement of an N -particle system (with $N \geq 3$) is much richer in comparison to the bipartite case. This is because quantum entanglement is a “monogamous” property that can be shared among the different parts in many non-equivalent ways and not freely, since the degree of entanglement between any two of its parts influences the degree of entanglement that can be shared with a third part [7, 8]. In the last two decades, there has been a great effort to recognize if a state with $N \geq 3$ is entangled and many analytic results have been found, see the review papers [9, 10]. However, even in the simplest case of three qubits, a general solution is still missing.

In this paper we tackle the still open question of finding

if a given state is entangled or separable as a classification problem.

Many classification problems can be solved efficiently by machine learning (ML) algorithms [11, 12], whose goal is to find a way to infer, from the available labeled data, the class of an unlabeled data point. Neural network ML algorithms for classification of entangled states have already been adopted in previous works, such as in Ref. [13] for the classification of two-qubit states labeled by means of Bell’s inequalities, or in Ref. [14], where a classifier is trained to recognize separable states of multi-qubit systems.

Another well-known ML algorithm is the Support Vector Machine (SVM) [15], that has been proved to be very solid in solving classification problems in image recognition [16], medicine [17, 18] and biophysics [19] among others. A collection of applications of the SVM algorithm can be found in the review paper [20]. The SVM approach has been recently investigated for classification of two-qubit states [21] and four-qubit states [22]. In particular, in the former paper, the authors use a SVM with exponential kernel, supported by a preliminary Convex Hull Approximation (CHA) analysis, to construct a separability-entanglement classifier via a supervised learning approach.

In this work, we put forward a new approach, exploiting the so-called kernel trick [12] with polynomial kernels, in which we train a SVM to classify entangled and separable states with a high predictive power, specializing the analysis first to the case of the two-qubit system, as a benchmark, and then generalizing it to the classification of different classes of entangled states of a three-qubit system. The performance of the algorithm is evaluated through the following metrics: the accuracy a , corresponding to the percentage of true positives, i.e. the (either separable or entangled) states that are correctly classified, the precision p , that measures the percentage of predicted entangled states that are really entangled, and the recall r , that measures the percentage of entan-

* claudio.sanavio@unibo.it

gled states that are correctly classified. The hyperparameters of the algorithm are chosen to get $p = 1$ and the maximum possible value of r on the validation set.

In this way we are effectively implementing an *entanglement witness* operator, which does not misclassifies separable states as entangled while also minimizing the number of entangled states that are misclassified as separable. Such witness operator is explicitly constructed for a linear kernel, obtaining a standard witness functional over the space of density matrices of the system, as well as in the case of a non-linear polynomial kernel of degree n . In the latter case, which corresponds to the construction of a witness functional over n -copies of the system, the performance of the algorithm increases. Finally, we present a digital quantum circuit which is able to implement the construction and the evaluation of such entanglement witnesses.

The content of the paper is as follows. In Section II we review the entanglement classification for two-qubit and three-qubit systems, with emphasis on the analytical results that we use to generate the dataset of pure and mixed states [9, 23–27]. In Section III we describe the SVM algorithm and the kernel trick. In Section IV we relate the separating hypersurface obtained via the SVM to the evaluation of an entanglement witness operator, providing for that also an operational procedure that can be implemented on a digital quantum computer.

In Section V we train the SVM with two-qubit states for which we have an exact classification, i.e. the PPT criterion, and eventually test our method on a large sample. Section VI is devoted to the extension of our analysis to three-qubit systems. Finally, in Section VII we draw our conclusion and give outlooks for possible future research directions.

II. ENTANGLEMENT CLASSIFICATION

In this section we give a brief overview of the entanglement in two-qubit and three-qubit systems, with emphasis on the different analytical results found in the literature to classify entangled and separable states.

Let us first consider a pure quantum state $|\psi\rangle$ defined in the Hilbert space $\mathcal{H}^{\otimes N}$ of N identical particles and described by the density operator $\hat{\rho} = |\psi\rangle\langle\psi|$.

The state $\hat{\rho}$ can be written as a product state of $M \leq N$ density operators [28] as

$$\hat{\rho} = \otimes_{j=1}^M \hat{\rho}_j \quad (1)$$

If $M = N$, the state is called fully separable, and each $\hat{\rho}_j$ acts on the one-particle space \mathcal{H} . If $M < N$, at least one of the density operators $\hat{\rho}_j$ describes more than one particle. If $M = 1$, the state $\hat{\rho}$ is called Genuinely Multipartite Entangled (GME).

In the case of a two-qubit state, with $N = 2$ and $\mathcal{H} = \mathbb{C}^2$, the system can either be separable, with $M = 2$, or entangled, with $M = 1$. In case the system is composed by three qubits, labeled as A, B and C, we have $N = 3$

and $\mathcal{H} = \mathbb{C}^2$. The state is fully separable (SEP) when $M = 3$, and therefore Eq.(1) reads as $\hat{\rho} = \hat{\rho}_A \otimes \hat{\rho}_B \otimes \hat{\rho}_C$. When $M = 2$ the state is called biseparable, and the density operator can be written as $\hat{\rho}_A \otimes \hat{\rho}_{BC}$ (A-BC), $\hat{\rho}_B \otimes \hat{\rho}_{AC}$ (B-AC) or $\hat{\rho}_C \otimes \hat{\rho}_{AB}$ (C-AB). When $M = 1$ the state is GME. GME states play an important role to devise robust distribution protocols [29].

Mixed states are a statistical mixture of pure states [1], and are described by a density operator of the form

$$\hat{\rho} = \sum_i p_i \hat{\rho}_i, \quad (2)$$

where $\sum_i p_i = 1$ and $\hat{\rho}_i$ are pure states. The mixed state $\hat{\rho}$ is fully separable when each pure state $\hat{\rho}_i$ is fully separable.

For a two-qubit system (as well as for a qubit-qutrit system), the PPT criterion [5, 6] offers a necessary and sufficient condition for labeling a quantum state as entangled: a two-qubit state is separable if and only if the partial transposition of the density matrix returns a density matrix. In all other cases, the PPT criterion offers only a necessary condition for separability [5].

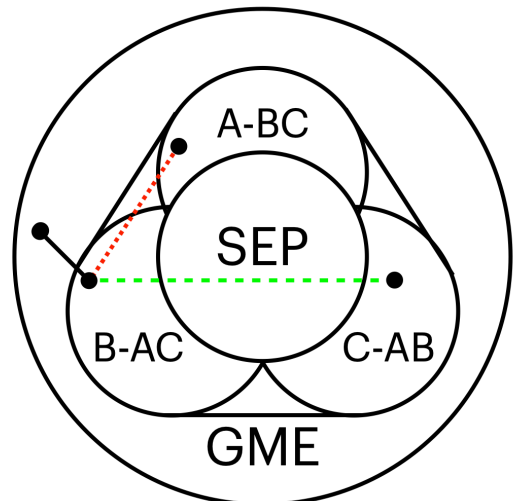


FIG. 1. The set of three-qubit states, called A, B and C, in terms of their level of entanglement. GME is the set of genuinely multipartite entangled states, SEP is the set of fully separable states, and A-BC, B-AC, C-AB represent the three possible factorization of the density operator in the biseparable case. Lines between sets of biseparable states represent the convex hull of biseparable states. The states are represented by circles. The segments joining them are their convex combinations. In the figure are represented convex combination of a state in B-AC with states in GME (black continuous line), in C-AB (green dashed line), and in A-BC (red dotted line). Depending on the weights given to the convex combination, the resulting state can lie in a different subset.

When considering the statistical mixture of states of three qubits, it is not straightforward to tell if the state is separable, biseparable or GME. Fig. 1 shows the general classification of three-qubits states. The convex set

of fully separable states is represented by the innermost circle, which is surrounded by the three kinds of biseparable states (A-BC, B-AC, C-AB) and finally by the GME states (largest circle). The segment joining two points represents the convex combination of two states. A statistical mixture of states can lie in any of the aforementioned classes. When the state is a statistical mixture of biseparable states, with respect to different partitions, the result can either be fully separable, or biseparable with respect to one of the partitions (green dashed segment in Fig.1), or biseparable with respect to none of the partitions (red dotted segment in Fig.1), thus forming the convex-hull of the biseparable states. It is known that all GME states can be obtained by applying SLOCC (stochastic local operations and classical communication) operations [23] either on the GHZ state

$$|\text{GHZ}\rangle = \frac{1}{\sqrt{2}}(|000\rangle + |111\rangle), \quad (3)$$

or on the W state

$$|W\rangle = \frac{1}{\sqrt{3}}(|100\rangle + |010\rangle + |001\rangle). \quad (4)$$

To correctly label as entangled or separable the three-qubit states used to construct the dataset for the ML classifier, we will resort to the analytical calculation of several entanglement measures [30]. In the remaining of this section we are going to introduce these analytical results, and we refer to Appendix A for the description of the states. In general, an entanglement measure μ is a real function defined on the set of the states that does not increase under the application of a SLOCC transformation [23], for which two states belong to the same class if you can transform one into the other with an invertible local operation, and the application of SLOCC brings a state into a class with lower or equal value of any entanglement measure. Pure states can be uniquely [23] labeled with the entanglement entropy S , the GME-concurrence C_{GME} and the three-tangle τ . The entanglement entropy S_m [31] corresponds to the Von-Neumann entropy calculated on the reduced system $\hat{\rho}_m = \text{Tr}_{ABC-m}[\hat{\rho}]$, with $m=A,B,C$:

$$S_m = -\text{Tr}[\hat{\rho}_m \ln \hat{\rho}_m]. \quad (5)$$

Genuinely Multipartite Entanglement concurrence [32] is defined as

$$C_{\text{GME}} = 2 \min_{i \in \mathcal{P}} \sqrt{1 - \text{Tr}[\hat{\rho}_i^2]}, \quad (6)$$

where $\mathcal{P} = \{AB, BC, AC\}$ is the set of the possible subsystems. Finally, the three-tangle τ [7] accounts for the genuinely multipartite entanglement of the GHZ class. Given the reduced density operator $\hat{\rho}_{AB}$ of the subsystem of the qubits A and B, we define the matrix $\tilde{\rho}_{AB} = (\sigma_y \otimes \sigma_y) \hat{\rho}_{AB} (\sigma_y \otimes \sigma_y)$, and we denote as $\lambda_{AB}^1, \lambda_{AB}^2$ the square root of the first and second eigenvalues of $\tilde{\rho}_{AB} \hat{\rho}_{AB}$. Then, the three-tangle is

$$\tau(\hat{\rho}) = 2(\lambda_{AB}^1 \lambda_{AB}^2 + \lambda_{AC}^1 \lambda_{AC}^2), \quad (7)$$

Class	Entanglement measure	label	
		SEP-vs-all	GME-vs-all
SEP	$S_m = 0$ for $m = A, B, C$	-1	-1
A-BC	$S_A = 0$ and $S_B, S_C > 0$	+1	-1
B-AC	$S_B = 0$ and $S_A, S_C > 0$	+1	-1
C-AB	$S_C = 0$ and $S_A, S_B > 0$	+1	-1
GHZ	$S_A, S_B, S_C > 0$ and $\tau > 0$	+1	+1
X-state	$\tilde{C}_{\text{GME}} > 0$	+1	+1
GHZ-sym	$\tilde{C}_{\text{GME}} > 0$ and $\tilde{\tau} > 0$	+1	+1
GHZ+W	$\tilde{\tau} > 0$	+1	+1

TABLE I. The classification of three-qubit states into entanglement classes. SEP are pure and mixed separable states. A-BC, B-AC, C-AB are the biseparable states. GME states divide into pure GHZ states, X-state, GHZ symmetric and statistical mixture of GHZ and W states, each labeled accordingly to the corresponding entanglement measures. The mixed states are classified using the CRE of the entanglement measures, denoted as \tilde{C}_{GME} and $\tilde{\tau}$. Fully separable states, pure and mixed, are labeled as -1 for both the SEP-vs-all and GME-vs-all classifiers, biseparable states are labeled as +1 in the SEP-vs-all classifier and -1 in the GME-vs-all classifier. GME states are always labeled as +1.

with λ_{AC}^i defined in a similar manner as λ_{AB}^i , for $i = 1, 2$. In the case of mixed states, the entanglement measures defined above fail to classify entangled states. One needs to calculate the Convex-Roof Extension (CRE) [33] of an entanglement measure μ defined as

$$\tilde{\mu}(\hat{\rho}) = \inf_P \sum_i p_i \mu(\hat{\rho}_i), \quad (8)$$

where P stands for all the possible decompositions $\{p_i, \hat{\rho}_i\}$ such that $\hat{\rho} = \sum_i p_i \hat{\rho}_i$, with $\sum_i p_i = 1$ and $\hat{\rho}_i$ is a pure state. This is in general difficult to calculate, and numerical solutions are usually very costly. Therefore, we use particular types of mixed states for which the analytic solutions for the CRE of the three-tangle [26, 27, 34], and the CRE of GME-concurrence [9, 24, 25] are available. These are the GHZ-symmetric states, the X-states, and the statistical mixture of GHZ and W states which are defined in Appendix A, which will be used in the following to construct the three-qubit dataset of the SVM algorithm. Table I summarizes how the labels are assigned to these different type of states.

III. SUPPORT VECTOR MACHINE

In this section we give a general introduction of the SVM.

Suppose that we have a dataset divided into two classes, composed of m pairs $\{(\mathbf{x}_i, y_i)\}$. The i -th observation $\mathbf{x}_i \in \mathbb{R}^d$ describes the characteristics of the point while $y_i = \pm 1$ labels its class. The SVM [12, 15] is a ML

classification algorithm based on the assumption that a hyperplane, or more generally a $d-1$ surface, which separates the two classes in the feature space exists. The hyperplane is defined through a linear decision function $f(\mathbf{x})$, with parameters $\mathbf{w} \in \mathbb{R}^d$, called weights, and $b \in \mathbb{R}$, called bias, given as

$$\begin{aligned} f(\mathbf{x}) &= \mathbf{w}^T \mathbf{x} + b, \\ \text{s.t.} \quad &\begin{cases} f(\mathbf{x}_i) > 0 & \text{if } y_i = +1, \\ f(\mathbf{x}_i) < 0 & \text{if } y_i = -1. \end{cases} \end{aligned} \quad (9)$$

The SVM algorithm aims to find the hyperplane (\mathbf{w}, b) that maximizes the margin $1/\|\mathbf{w}\|$ ($\|\cdot\|$ being the euclidean norm) between itself and the closest points of the two classes. Fig. 2 shows a two-dimensional dataset made of two classes, the circles and the triangles, that are separated by a hyperplane.

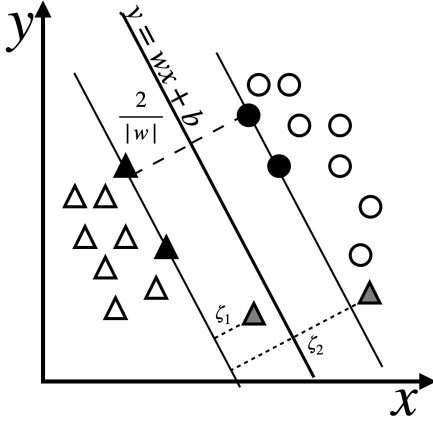


FIG. 2. Representation of the separating line between two data classes, the circles and the triangles, in a two-dimensional plane. The two classes are linearly separable by the straight line $y = wx + b$. The margin, i.e. the distance between the hyperplane and the closest data point of each class is $1/|w|$. The lines $y - wx - b = \pm 1$ cross the support vectors, here drawn with a filled shape. The slack variables ζ_1 and ζ_2 penalize the possible misclassification of the two gray triangle points.

Maximizing the margin corresponds to finding

$$\begin{aligned} \min_{\mathbf{w}} \quad & \frac{1}{2} \|\mathbf{w}\|^2, \\ \text{s.t.} \quad & y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1, \text{ for } i = 1, \dots, m. \end{aligned} \quad (10)$$

We can assume that the hyperplane defined by the decision function f is not able to perfectly separate the two classes. Thus, we can introduce m slack variables ζ_i , one for each observation \mathbf{x}_i , to account for the constraints of f in Eq. (9) that are not satisfied. Hence, Eq.(10) is

refined into

$$\min_{\mathbf{w}, \zeta} \quad \frac{1}{2} \|\mathbf{w}\|^2 + \sum_{i=1}^m \lambda_{y_i} \zeta_i, \quad (11a)$$

$$\text{s.t.} \quad y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1 - \zeta_i, \quad (11b)$$

$$\zeta_i \geq 0. \quad (11c)$$

The constants $\lambda_{y_i} \geq 0$ are called regularization parameters. Geometrically, the action of the slack variables ζ_i can be seen as a non-smooth deviation of the hyperplane, as shown in Fig. 2.

In order to construct a Lagrange function from which we can variationally derive Eq. (11a), we use the so called Karush-Kuhn-Tucker (KKT) complementary conditions [12]: after introducing dual variables α_i and λ_{y_i} , the constraints can be re-expressed as:

$$\alpha_i [y_i(\mathbf{w}_i^T \mathbf{x}_i + b) - 1 + \zeta_i] = 0, \quad (12a)$$

$$\zeta_i (\alpha_i - \lambda_{y_i}) = 0. \quad (12b)$$

These KKT conditions state that if the variable \mathbf{x}_i lies on the correct side of the hyperplane, with $y_i(\mathbf{w}_i^T \mathbf{x}_i + b) > 1$, then $\alpha_i = 0$ and $\zeta_i = 0$. On the other hand, if the variable \mathbf{x}_i lies exactly on the margin, with $y_i(\mathbf{w}_i^T \mathbf{x}_i + b) = 1$, then $\alpha_i = \lambda_{y_i}$, and $\zeta_i = 0$. Lastly, if the point \mathbf{x}_i lies on the wrong side of the hyperplane, we have $0 < \alpha_i < \lambda_{y_i}$ and $\zeta_i > 0$. In general, a number $m' < m$ of points \mathbf{x}_i have a corresponding dual variable $\alpha_i \neq 0$. These points, that lie on the separating hypersurface, are called support vectors.

The KKT conditions allow us to write the Lagrange function of the problem as

$$\begin{aligned} \mathcal{L} = & \frac{1}{2} \|\mathbf{w}\|^2 + \sum_{i=1}^m \alpha_i [1 - \zeta_i - y_i(\mathbf{w}^T \mathbf{x}_i + b)] \\ & + \sum_{i=1}^m \lambda_{y_i} \zeta_i - \sum_{i=1}^m \beta_i \zeta_i, \end{aligned} \quad (13)$$

where $\alpha_i, \beta_i \geq 0$ are the Lagrange multipliers introduced for the constraints (11b) and (11c) respectively. The optimality condition is satisfied when the partial derivatives with respect to the primal variables \mathbf{w}, b, ζ_i vanish,

$$\partial_{\mathbf{w}} \mathcal{L} = \mathbf{w} - \sum_{i=1}^m y_i \alpha_i \mathbf{x}_i = 0 \quad (14a)$$

$$\partial_b \mathcal{L} = - \sum_{i=1}^m y_i \alpha_i = 0 \quad (14b)$$

$$\partial_{\zeta_i} \mathcal{L} = \lambda_{y_i} - \alpha_i - \beta_i = 0, \forall i. \quad (14c)$$

Substituting Eqs. (14a),(14b) and (14c) into Eq. (13) yields the Lagrange function expressed in terms of the multipliers $\alpha = (\alpha_1, \dots, \alpha_m)$

$$\mathcal{L}(\alpha) = \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i,j=1}^m \alpha_i y_i K_{ij} y_j \alpha_j, \quad (15)$$

where K is called kernel matrix, and the component K_{ij} corresponds to the inner product $\langle \mathbf{x}_i, \mathbf{x}_j \rangle$ between the feature vectors.

The matrix K encodes the power of the SVM algorithm, that is able to perform a separation with a higher-degree decision function. In principle, to do so one has to embed the feature vectors into a higher dimensional space with an embedding function $\Phi : \mathbb{R}^d \rightarrow \mathbb{R}^D$, with $D > d$, and find the linear hyperplane that separates the data classes in the embedding space \mathbb{R}^D . As a result, the Lagrange function can be written as in Eq.(15), with $K_{ij} = \langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j) \rangle$ being the inner product in the embedding space. The procedure of adopting a specific kernel matrix K in order to move to a higher-dimensional space is known as kernel trick [12].

One common choice is the polynomial kernel of degree n , with $K(\mathbf{x}, \mathbf{x}') = (\langle \mathbf{x}, \mathbf{x}' \rangle + 1)^n$. It corresponds to embedding the d dimensional vectors \mathbf{x} and \mathbf{x}' into a space of dimension [12]

$$D = \binom{n+d-1}{n} \quad (16)$$

where we can look for a separating $(D-1)$ -surface. The components of $\Phi(\mathbf{x})$ in the D -dimensional space come from the D terms in the multinomial expansion of degree n . The constant 1 plays an important role, as it allows the presence of all the terms up to degree n in the decision function f , while in its absence the decision function has only the terms of degree exactly n . As the value of D is extremely large for real case scenarios, the kernel trick saves us from having to explicitly embed the data in a much larger space and offers a smart way of coping with the embedding process.

The classification of a new data point \mathbf{x} consists simply in calculating the decision function

$$f(\mathbf{x}) = \sum_{i=1}^{m'} \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + b, \quad (17)$$

and the label will be given by the sign of $f(x)$. Note that the sum is made only over the m' support vectors, as for the other points the Lagrange multipliers are zero.

IV. SVM-DERIVED ENTANGLEMENT WITNESSES

In this section we show how the separating hyperplane provided by the SVM algorithm as a result of the minimization in Eq. (11a) can be translated into a Hermitian operator that we can measure to detect entanglement. We also describe a procedure that can be implemented on a quantum computer in order to measure its mean value. The hyperplane inferred by the SVM has a huge similarity with the concept of entanglement witness. The classification in the SVM is calculated by considering the

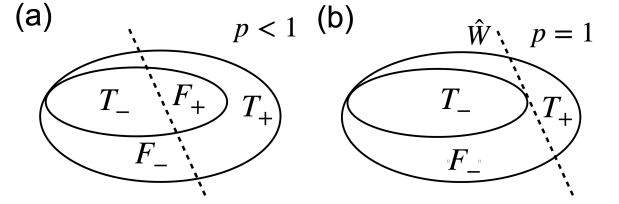


FIG. 3. Schematic representation of a classification performed on a convex set. (a) the separating hyperplane, represented by the dotted line, divides the two classes +1 on the right, and -1 on the left, with a certain amount of true and false positives, T_+ , F_+ , and true and false negatives T_- , F_- , therefore the precision $p < 1$. (b) the separating hyperplane has been built to mimic the behavior of an entanglement witness operator \hat{W} . In this case, there are not false positives, and the precision $p = 1$.

sign of the decision function $f(x)$. In a similar manner, an entanglement witness operator \hat{W} is such that if $\text{Tr}[\hat{W}\hat{\rho}] > 0$ then the state $\hat{\rho}$ is entangled [28], providing a sufficient but not necessary condition for entanglement detection. The convexity of separable states ensures us that we can find such operator [1].

Consistently with the choices of Table I, we assign to the separable states the label -1 and to the entangled states the label +1. In order to evaluate the performance of the classifier, we use the accuracy

$$a = (T_+ + T_-)/m, \quad (18)$$

where m is the total number of states considered in the dataset. Here T_+ , T_- are the number of true positives/negatives respectively, i.e. the number of entangled/separable states correctly classified as entangled/separable. Furthermore, the precision p and the recall r associated to the classification of entangled states are given by

$$p = T_+/(T_+ + F_+), \quad (19)$$

$$r = T_+/(T_+ + F_-), \quad (20)$$

where F_+ , F_- represent the number of false positives/negatives respectively, i.e. the number of separable/entangled states misclassified as entangled/separable. The condition $p = 1$ indicates that we are not misclassifying separable states as entangled. Therefore, the hyperplane (\mathbf{w}, b) that we find with this condition respects the definition of entanglement witness, see Fig. 3. We can get this condition by fine-tuning the regularization hyperparameters λ_+ and λ_- defined in equation (13). In fact, we assign to each data point the regularization parameter λ_+ or λ_- depending on whether it belongs to the class of entangled or separable states, respectively. The condition $\lambda_+ > \lambda_-$ penalizes the misclassification of entangled states. On the other hand, the condition $\lambda_- > \lambda_+$ penalizes the misclassification of separable states, favoring thus the classifier to correctly predict -1, leading in turn to a

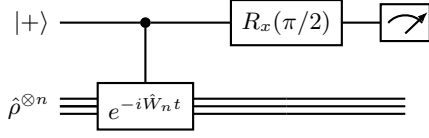


FIG. 4. Scheme for measuring the mean value of the operator \hat{W}_n . The control qubit is in the state $|+\rangle$. A controlled unitary gate with operator $\exp(-i\hat{W}_n t)$ is applied on the n copies of the system state $\hat{\rho}$. The measurement is performed on the control qubit after a $\pi/2$ rotation R_x . Thus, the mean value is obtained as $\langle \hat{W}_n \rangle = \frac{p_1 - p_0}{2t}$, with p_0 and p_1 being the probability that the outcome of the measurement is 0 or 1 respectively.

higher precision in classification of entangled states at the cost of a lower recall. In terms of the classification algorithm, the witness operator \hat{W} is related to a hyperplane with perfect precision of entangled states. When the number of false positive F_+ is minimized, we mimic an entanglement witness operator \hat{W} , with the property that $\text{Tr}[\hat{W}\hat{\rho}] > -b$ if and only if the state is entangled, with b being the bias of the hyperplane, see Eq. (9).

The operation $\text{Tr}[\hat{W}\hat{\rho}]$ is related to a SVM hyperplane with linear kernel, $K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T \mathbf{x}_j$. In the following, we will show that a more precise classification of entangled states can be carried out by non-linear kernels of degree n . The latter correspond to exotic witness operators \hat{W}_n that work on n copies of the state $\hat{\rho}$. With little algebra one can show that the decision function (17) can be written as

$$f_n(\hat{\rho}) = \text{Tr}[\hat{W}_n \hat{\rho}^{\otimes n}] + b \quad (21)$$

with the non-linear witness operator \hat{W}_n given by

$$\hat{W}_n = \sum_{i=1}^{m'} \sum_{l=1}^n y_i \alpha_{i,l} (\hat{\omega}_i^{\otimes l} \otimes \mathbb{1}_2^{(n-l)}) \quad (22)$$

where m' is the number of support vectors states. Here, $\mathbb{1}_2^{(k)}$ indicates the identity matrix in the space of k qubits, with dimension 2^k . The coefficients of the operator $\hat{\omega}_i$ depend only on the i -th support vector state, which can be found, together with the values of $\alpha_{i,l}$, via a classical training of the SVM classifier.

In this way, the classification of an entangled state has become the evaluation of the mean value of the Hermitian operator \hat{W}_n on the state $\hat{\rho}^{\otimes n}$. As far as we know, Eq. (22) is a novel equation that relates a classical SVM hyperplane to a non-linear witness operator.

In order to measure the mean value of the Hermitian operator \hat{W}_n we can use the procedure described in [35] and shown in Fig. 4, where a controlled unitary operator $C-U$, with $U = \exp(-i\hat{W}_n t)$, is applied on a system composed by the n copies of the system $\hat{\rho}$.

The control qubit is initialized in the state $|+\rangle$. For small t , we can expand the operator $U \approx 1 - it\hat{W}_n +$

$\mathcal{O}(t^2)$. Measuring the mean value of the Pauli matrix $\langle \sigma_y \rangle$ on the control qubit corresponds to measuring $t\langle \hat{W}_n \rangle$. This can be achieved by transforming the σ_y basis into the computational basis with the operator $R_x(\pi/2) = \exp(-i\sigma_x \pi/4)$, and measuring the control qubit. The probability of getting as outcome the state $|0\rangle$ is $p_0 = 1/2 - t\langle \hat{W}_n \rangle$, whereas the probability of getting the outcome $|1\rangle$ is $p_1 = 1/2 + t\langle \hat{W}_n \rangle$. Calculating $p_1 - p_0$ and dividing by $2t$ leads to the wanted result. Notice that the operator U is not a local operation, and therefore, the state $\hat{\rho}$ after the protocol will not in general conserve the level of entanglement. Thus, if the goal of the circuit is to use an entangled state as input of a quantum protocol, we would need $n + 1$ copies of $\hat{\rho}$.

V. TWO-QUBIT SYSTEM

In this section we show the results of the SVM classification for a two-qubit system.

In order to provide the SVM a suitable dataset, we need to uniformly sample the density matrices with respect to the Hilbert-Schmidt (HS) measure [36]. Thus, we generate 4×4 matrices A with \mathbb{C} -numbers elements $A_{ij} \in \mathcal{N}(0, 1)$, $\mathcal{N}(0, 1)$ being the normal distribution with zero mean and unitary variance. The quantum state is obtained from the positive trace-one Hermitian matrix $\hat{\rho} = AA^\dagger / \text{Tr}[AA^\dagger]$. Uniformly generating two-qubit quantum states accordingly to the HS measure leads to an imbalance in the classes, as the ratio between separable and entangled states is approximately 1 to 4 (≈ 0.24) [37]. We balance our dataset in order to get 50% of entangled and 50% of separable states. This inter-class balance allows us to have more control of the hyperparameters. Indeed, in this case, when $\lambda_- = \lambda_+$ the SVM does not prefer any labeling on a new data point. Hence, we separately generate an equal number of entangled and separable states, labeling them with the PPT criterion [6]. Our dataset consists of 10^6 samples for each class, 98% of which are used in the training set, 1% are used in the validation set, and 1% are used in the test set. Each data point lives in the feature space \mathbb{R}^{15} , and is obtained by taking the real and imaginary part of the components of the 4×4 density matrix, which is hermitian and trace-one.

In order to fine-tune the hyperparameters, we look at the validation set. The accuracy (18), the precision and the recall (20) depend on the ratio λ_-/λ_+ . Fig. (5) shows the obtained precisions and recalls for different degrees n of the kernel, $n = 2$ (a), $n = 3$ (b), $n = 4$ (c), and the resulting accuracies on the validation set, calculated varying the hyperparameters λ_\pm . Note that using the imbalanced training set, the result would have been analogous, but shifted, and we would have got the same results using a larger value of λ_+ . For each degree of the polynomial kernel n , we find with a grid search the value of λ_-/λ_+ s.t. on the validation set $p = 1$ and r is maximum. This point, with $\lambda_- = \lambda_n^*$ and $\lambda_+ = 1$, is located

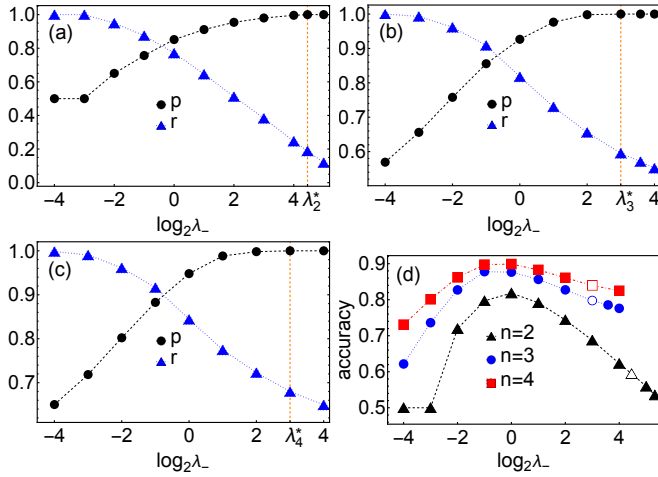


FIG. 5. Two qubits. Precision p and recall r defined in Eqs. (19) and (20) of the two-qubits entanglement classifier as a function of the regularization parameter λ_- ($\lambda_+ = 1$), on a validation set of 10^4 samples for each class in the case of polynomial kernel of degree (a) $n = 2$, (b) $n = 3$, and (c) $n = 4$. The vertical dotted line refers to the point $\lambda_- = \lambda_n^*$ for each $n = 2, 3, 4$, where r is maximum with the condition $p = 1$ satisfied. (d) The accuracy on the validation set for the classifier different degrees $n = 2, 3, 4$. The empty markers correspond to $\lambda_- = \lambda_n^*$.

at the orange vertical line in the plots in Figs. 5(a),(b) and (c). Fig. 5(d) shows the total accuracy evaluated on the validation set. We see that the classification accuracy increases for higher degree n . Note that the highest accuracy is reached in correspondence of the point $\lambda_+ = \lambda_-$, corresponding to a balanced dataset with an equal number of entangled and separable states. In fact, the value λ_-/λ_+ where we reach highest accuracy changes depending on the population of the two classes in the dataset.

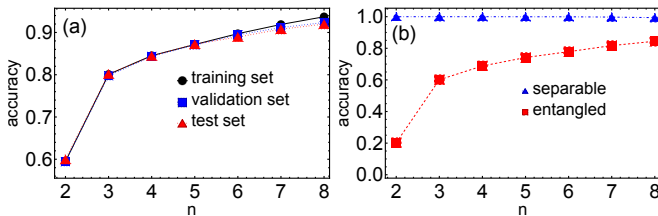


FIG. 6. Two qubits. (a) The accuracy obtained by the n degree polynomial classifier of two-qubit states, with $n = 1, 2, \dots, 8$. The classifier gains accuracy with a higher n . The trend of the curve is the same for the three sets, showing that the classifier is able to generalize. (b) The accuracy of the classifier on the two classes.

Different degrees n of the polynomial kernel lead to different performances of the algorithm. Fig. 6(a) shows the accuracy of the classifier with $\lambda_+ = 1$ and $\lambda_- = \lambda_n^*$ on training, validation, and test set for $n = 1, 2, \dots, 8$. We see that by increasing the degree n we get a better accuracy in classification on the test set, reaching the

value $a \approx 0.91$ for the polynomial SVM with degree 8, which indicates that the classifier is capable of generalization. From Fig. 6(b) we see that accuracy is almost one for separable states, meaning that we selected the hyperparameters such that separable states are almost always correctly classified, while it tends to a value slightly higher than 0.8 for entangled states.

VI. THREE-QUBIT SYSTEM

In the same spirit as the two-qubits classification, we look for the hyperplane that minimizes the number of misclassified separable states. The goal is to be as close as possible to the border of the convex set of the separable states and to the definition of entanglement witness. However, the three-qubit classification of entangled state has a different inherent nature, as we aim to classify our state into three different classes: the class of fully separable states, the class of biseparable states and the class of GME states. This is achieved by introducing two classifiers. We call SEP-vs-all the classifier that aims at splitting the space into fully separable states and either biseparable or GME states; this classifier defines the operator witness \hat{W}_S . We call GME-vs-all the classifier that aims at splitting the GME states from the set of biseparable and fully separable states; this corresponds to the operator witness \hat{W}_E . Note that the classifiers here defined are able to distinguish only the GME states that belong to the classes that we have used in the dataset, see Table I. In fact, they are not able to classify states whose classes have not appeared in the training set. This is intrinsic in the nature of ML classifiers, that rely on known data to label new states.

For a three-qubit system, the feature space is composed of 63 real numbers describing the real and imaginary parts of the independent components of the 8×8 hermitian and trace-one density matrix. The datasets of the two classifiers are obtained generating 2×10^5 states, 90% of which are used in the training set, 5% are used in the validation set, and 5% are used in the test set. Pure and mixed fully separable states are easily generated by construction and labeled as -1 for both the datasets of the two classifiers. In the SEP-vs-all classifier, mixed fully separable states are the 50% of the dataset, whereas they are 25% of the dataset used for the GME-vs-all classifier.

GME states are labeled as $+1$ for both SEP-vs-all and GME-vs-all classifiers.

The GME states are labeled according to the three-tangle for generated pure GHZ states [23], defined in Eq. (3), GHZ-symmetric states [34], and statistical mixture of GHZ, W , and \bar{W} states [26]. In the case of mixed X-state [27] we use the GME-concurrence, Eq. (6). We refer the reader to Appendix A for the definition of these states. Once we have generated these states, a random local unitary transformation was applied on each qubit to increase the number of data and fill the different com-

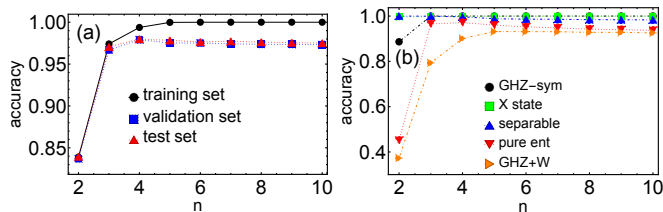


FIG. 7. Three qubits. The accuracy obtained by the GME-vs-all classifier. (a) The accuracy on the training, validation and test sets. (b) The accuracy for different entanglement-type states used in the training set, as a function of the kernel degree n . The results are shown at the fixed ratio λ_-/λ_+ obtained maximizing r on the validation set with the constraint $p = 1$.

ponents of the density operator. Each category of GME states is generated with the same number of samples, and they are 25% of the SEP-vs-all dataset, and 50% of the GME-vs-all dataset. The qualitative difference in the datasets of the two classifiers lies in the label of biseparable states, as given in Table I. For the classifier SEP-vs-all, only pure biseparable states can be considered in the dataset, as a mixture of biseparable state can also be fully separable. These states are generated and labeled as +1. Their category represent the 25% of this dataset. On the other hand, the dataset of classifier GME-vs-all contains mixture of biseparable states with respect to the same partition, and labeled as -1. This is because mixture of biseparable states with respect to different partitions can either be biseparable or fully separable, as shown in Fig 1. They are the 25% of the samples in the GME-vs-all dataset.

Let us examine the accuracy of both classifiers on the different classes of states that we considered.

Fig. 7(a) shows the accuracy on the test set of the classifiers GME-vs-all for different values of the degree n , $n = 2, 3, \dots, 10$. Fig. 8(a) shows the accuracy obtained for the classifier SEP-vs-all on the test set. Figs. 7(b) and 8(b) show the classification accuracy for each category of entanglement states used in the dataset on the corresponding test set. We notice that GHZ+W states are more difficult to classify for both GME-vs-all and SEP-vs-all classifiers. Instead, GHZ symmetric states are correctly labeled with an accuracy ≈ 1 by the GME-vs-all classifier, and with accuracy > 0.9 for $n \geq 3$ by the SEP-vs-all classifier.

Contrary to what happens for the two-qubit classifier, we observe that the GME-vs-all classifier shows overfitting for $n > 4$. This feature is captured by the spread between the accuracies of the training and the validation sets shown in Fig. 7(a). The same happens to the SEP-vs-all classifier for $n > 6$.

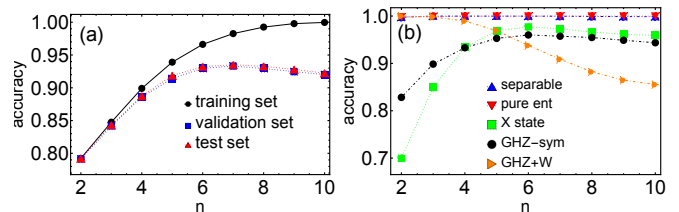


FIG. 8. Three qubits. The accuracy obtained by the SEP-vs-all classifier. (a) The accuracy on the training, validation and test sets. (b) The accuracy for different entanglement-type states used in the training set, as a function of the kernel degree n . The results are shown at the fixed ratio λ_-/λ_+ obtained maximizing r on the validation set with the constraint $p = 1$.

VII. CONCLUSION

In this work we have investigated the possibility of using a SVM for classification of entangled states, for two-qubit and three-qubit systems.

The SVM algorithm, which is known to be suitable for binary classification of data points belonging to convex sets, learns the best hyperplane that separates the set of separable states from the set of entangled states. This construction has a strong analogy with that of an entanglement witness operator, for which the sign of the mean value calculated over the state, is related to the separability of the latter. However, witness operators are usually linear operators. In this paper we have shown that, exploiting the SVM algorithm and the kernel trick, we can use many copies of the state and construct non-linear witness operators, which show an increasing classification accuracy. We believe that these constructions can help us to gain knowledge about the geometry of the set of entangled states.

In the application of the SVM to the two-qubit system, we have evaluated accuracy, precision and recall for kernels of degree n up to 8, reaching an increasing accuracy. For three-qubit systems we have trained two classifiers, named SEP-vs-all, that distinguishes fully separable states to all the others, and GME-vs-all, for detection of GME states, with kernels of degree n up to 10. The SEP-vs-all classifier shows an accuracy on the test set of 92.5%, whereas the GME-vs-all classifier reaches an accuracy of 98%, in the best case corresponding to $n = 4$. Furthermore, we have checked the accuracy among the different types of states that we have used in the dataset.

Overall, the SVM algorithm has shown a high accuracy in detecting entanglement. We have also discussed how the classifier can be trained to favor higher precision in recognition of entangled states, at the expenses of a lower global accuracy. This feature can be used in quantum computation algorithms, where entanglement plays an important role for quantum advantage. Moreover, if a quantum protocol needs a specific type of entangled state, the SVM algorithm can be trained to recognize it and use it as input.

As entanglement classification is still an open line of research both for the theoretical and fundamental aspects, we believe machine learning algorithms can provide a useful tool of analysis.

ACKNOWLEDGMENTS

We thank T. Apollaro and M. Consiglio for useful discussions. This research is funded by the International Foundation Big Data and Artificial Intelligence for Human Development (IFAB) through the project “Quantum Computing for Applications”. E. E. is partially supported by INFN through the project “QUANTUM” and the QuantERA 2020 Project “QuantHEP”.

Appendix A: Three-qubit entangled mixed states

In this appendix we briefly describe the three-qubit GME states that we generated for the dataset of the SEP-vs-all and GME-vs-all classifiers.

A GHZ symmetric state is characterized by three fundamental properties, i) it is invariant under index permutation of the qubits, ii) it is invariant under simultaneous spin flip of the qubits, and iii) it is invariant under the rotation

$$U(\phi_1, \phi_2) = e^{i\phi_1 \hat{\sigma}_z} \otimes e^{i\phi_2 \hat{\sigma}_z} \otimes e^{-i(\phi_1 + \phi_2) \hat{\sigma}_z}. \quad (\text{A1})$$

This state can be written as a mixture of the two GHZ

states, with density operators $\hat{\rho}_{\text{GHZ}\pm}$, and of the maximally mixed state with density operator $1/8$,

$$p \hat{\rho}_{\text{GHZ}+} + q \hat{\rho}_{\text{GHZ}-} + (1 - p - q) 1/8.$$

When $q = 0$, we have the so-called Werner state [38], with a GHZ state combined with the maximally mixed state. In Refs. [9, 24, 25], Eltschka et al. gave a complete classification of GHZ symmetric states in terms of the relation between the coefficients p and q .

The X-state owes its name to the shape of the density matrix, which has non null values on the diagonal and anti-diagonal elements. Thus the X-state density operator can be written as the sum $\hat{D} + \hat{A}$, with $\hat{D} = \text{diag}(a_1, a_2, \dots, a_n, b_n, \dots, b_1)$ the $2n \times 2n$ diagonal matrix, and $\hat{A} = \text{anti-diag}(z_1, \dots, z_n, z_n^*, \dots, z_1^*)$ being an anti-diagonal matrix. The conditions $\text{Tr}[\hat{D}] = 1$ and $\hat{D} \geq \hat{A}$ ensure that the operator $\hat{D} + \hat{A}$ describes a physical state. The GME-concurrence of the X-state can be analytically calculated [27], and when its value is positive, the X-state belongs to the GME class. Note that the GHZ symmetric states are particular X-states with $z_i = 0$ for $i = 2, \dots, n$.

Another class that we have used to train our classifier is the statistical mixture of GHZ, W, and the flipped W state $\hat{W} = \frac{1}{\sqrt{3}}(|011\rangle + |101\rangle + |110\rangle)$, that we have labeled as GHZ+W, with density operator

$$p \hat{\rho}_{\text{GHZ}} + q \hat{\rho}_{\text{W}} + (1 - p - q) \hat{\rho}_{\bar{W}}.$$

The analytic calculation of the three-tangle measure for those states has been given in [24, 26, 34].

-
- [1] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki. Quantum entanglement. *Rev. Mod. Phys.*, 81:865–942, Jun 2009.
 - [2] R. Jozsa and N. Linden. On the role of entanglement in quantum-computational speed-up. *Proc. R. Soc. London A*, 459(2036):2011–2032. Publisher: Royal Society.
 - [3] W. K. Wootters and W. S. Leng. Quantum entanglement as a quantifiable resource [and discussion]. *Phil. Trans R. Soc. A*, 356(1743):1717–1731. Publisher: The Royal Society.
 - [4] O. Gühne and G. Tóth. Entanglement detection. *Phys. Rep.*, 474(1):1–75.
 - [5] A. Peres. Separability criterion for density matrices. *Phys. Rev. Lett.*, 77:1413–1415, Aug 1996.
 - [6] M. Horodecki, P. Horodecki, and R. Horodecki. Separability of mixed states: Necessary and sufficient conditions. *Phys. Lett. A*, 223(1):1–8.
 - [7] V. Coffman, J. Kundu, and W. K. Wootters. Distributed entanglement. *Phys. Rev. A*, 61:052306, Apr 2000.
 - [8] T. J. Osborne and F. Verstraete. General monogamy inequality for bipartite qubit entanglement. *Phys. Rev. Lett.*, 96:220503, Jun 2006.
 - [9] C. Eltschka and J. Siewert. Quantifying entanglement resources. *J. Phys. A: Math. Theor.*, 47(42):424005.
 - [10] S. Szalay. Multipartite entanglement measures. *Phys. Rev. A*, 92(4):042329.
 - [11] S. B. Kotsiantis, I. D. Zaharakis, and P. E. Pintelas. Machine Learning: a review of classification and combining techniques. *Artif. Intell. Rev.*, 26(3):159–190.
 - [12] B. Schölkopf and A. J. Smola. *Learning with kernels: support vector machines, regularization, optimization, and beyond*. Adaptive computation and machine learning. MIT Press.
 - [13] Y. C. Ma and M. H. Yung. Transforming Bell’s inequalities into state classifiers with Machine Learning. *Npj Quantum Inf.*, 4(1):1–10. Number: 1 Publisher: Nature Publishing Group.
 - [14] C. Harney, S. Pirandola, A. Ferraro, and M. Paternostro. Entanglement classification via neural network quantum states. *New J. Phys.*, 22(4):045001.
 - [15] B. E. Boser, I. M. Guyon, and V. N. Vapnik. A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory, COLT ’92*, pages 144–152. Association for Computing Machinery.
 - [16] M. A. Chandra and S. S. Bedi. Survey on SVM and their application in imageclassification. *Int. J. Inf. Technol.*, 13(5):1–11.
 - [17] A. Hazra, S. K. Mandal, and A. Gupta. Study and analysis of breast cancer cell detection using naïve Bayes,

- SVM and ensemble algorithms. *Int. J. Comput. Appl.*, 145(2):39–45, Jul 2016.
- [18] V. J. Kadam, S. S. Yadav, and S. M. Jadhav. Soft-margin SVM incorporating feature selection using improved elitist GA for arrhythmia classification. In Ajith Abraham, Aswani Kumar Cherukuri, Patricia Melin, and Niketa Gandhi, editors, *Intelligent Systems Design and Applications*, pages 965–976. Springer International Publishing.
- [19] Y. D. Cai, X. J. Liu, X. B. Xu, and G. P. Zhou. Support vector machines for predicting protein structural class. *BMC Bioinformatics*, 2(1):3.
- [20] J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua, and A. Lopez. A comprehensive survey on support vector machine classification: Applications, challenges and trends. *Neurocomputing*, 408:189–215.
- [21] S. Lu, S. Huang, K. Li, J. Li, J. Chen, D. Lu, Z. Ji, Y. Shen, D. Zhou, and B. Zeng. Separability-entanglement classifier via machine learning. *Phys. Rev. A*, 98:012315, Jul 2018.
- [22] S. V. Vintskevich, N. Bao, A. Nomerotski, P. Stankus, and D. A. Grigoriev. Classification of four-qubit entangled states via machine learning, 2022.
- [23] W. Dür, G. Vidal, and J. I. Cirac. Three qubits can be entangled in two inequivalent ways. *Phys. Rev. A*, 62(6):062314.
- [24] C. Eltschka, A. Osterloh, J. Siewert, and A. Uhlmann. Three-tangle for mixtures of generalized GHZ and generalized w states. *New J. Phys.*, 10(4):043014, apr 2008.
- [25] C. Eltschka and J. Siewert. A quantitative witness for greenberger-horne-zeilinger entanglement. *Sci. Rep.*, 2(1):942.
- [26] E. Jung, M. R. Hwang, D. Park, and J. W. Son. Three-tangle for rank-three mixed states: Mixture of greenberger-horne-zeilinger, w, and flipped-w states. *Phys. Rev. A*, 79(2):024306.
- [27] S. M. Hashemi Rafsanjani, M. Huber, C. J. Broadbent, and J. H. Eberly. Genuinely multipartite concurrence of n -qubit x matrices. *Phys. Rev. A*, 86:062303, Dec 2012.
- [28] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press.
- [29] T. J. G. Apollaro, C. Sanavio, W.J. Chetcuti, and S. Lorenzo. Multipartite entanglement transfer in spin chains. *Phys. Lett. A*, 384(15):126306.
- [30] M. B. Plenio and S. Virmani. An introduction to entanglement measures. *Quantum Inf. Comput.*, 7(1):1–51.
- [31] V. Vedral. The role of relative entropy in quantum information theory. *Rev. Mod. Phys.*, 74:197–234, Mar 2002.
- [32] W. K. Wootters. Entanglement of formation of an arbitrary state of two qubits. *Phys. Rev. Lett.*, 80:2245–2248, Mar 1998.
- [33] A. Uhlmann. Entropy and optimal decompositions of states relative to a maximal commutative subalgebra. *Open Syst. Inf. Dyn.*, 5(3):209–228.
- [34] R. Lohmayer, A. Osterloh, J. Siewert, and A. Uhlmann. Entangled three-qubit states without concurrence and three-tangle. *Phys. Rev. Lett.*, 97:260502, Dec 2006.
- [35] F. Tacchino, A. Chiesa, S. Carretta, and D. Gerace. Quantum computers as universal quantum simulators: State-of-the-art and perspectives. *Adv. Quantum Technol.*, 3:1900052.
- [36] K. Zyczkowski and H. J. Sommers. Induced measures in the space of mixed quantum states. *J. Phys. A: Math. Theor.*, 34(35):7111–7125, Aug 2001.
- [37] P. B. Slater. Dyson indices and hilbert–schmidt separability functions and probabilities. *J. Phys. A: Math. Theor.*, 40(47):14279–14308, nov 2007.
- [38] R. F. Werner. Quantum states with Einstein-Podolsky-Rosen correlations admitting a hidden-variable model. *Phys. Rev. A*, 40:4277–4281, Oct 1989.