Quantum Machine Intelligence (2025) 7:112
https://doi.org/10.1007/s42484-025-00339-4

RESEARCH ARTICLE

Classification of quantum correlations via quantum-inspired machine learning

Giuseppe Sergioli[1] · Andrés Camilo Granda Arango[1] · Carlo Cuccu[1] · Stefania Centrone[3] · Carla Rieger[4,5] · Roberto Giuntini[1,2]

Received: 7 January 2025 / Accepted: 7 November 2025

Abstract
Quantum information theory, and in particular, the theory of quantum state discrimination, has enabled the development of a supervised multi-class classification algorithm. Inspired by the *Pretty Good Measurement* (PGM), a quantum-inspired classifier named the PGM Classifier has been designed, capable of classifying among multiple classes without resorting to the One versus One or One versus Rest strategy. In the recent past, several approaches of quantum machine learning have been introduced to classify entangled and separable states. In this article, we apply the PGM Classifier to discriminate among factorized states, separable states, and entangled states. We analyze from the simplest case (**2** qubits) up to a system of (**5** qubits) and offer a comparison between our quantum-inspired classifier and other classical classifiers. Finally, we propose a refined classification of non-locality based on the violation of the Mermin and Svetlichny inequalities.

Keywords Quantum-inspired machine learning · PGM Classifier · Quantum entanglement · Quantum non-locality.

1 Introduction

The rapidly growing field of quantum machine learning (QML) lies at the intersection of quantum information science and machine learning, offering new perspectives and methodologies for understanding and leveraging quantum correlations. This interdisciplinary approach aims to transcend the limitations of traditional computational methods, leveraging the inherent advantages of quantum systems for more sophisticated data analysis and processing.

Quantum entanglement, a fundamental aspect in quantum computing and information theory, presents both a challenge and an opportunity for advancing computational capabilities.

Machine learning tools, including support vector machine (SVM) and deep convolutional neural networks, have been applied with promising results in the classification and detection of entanglement across multi-qubit systems (El Ayachi and El Baz 2022; Pawłowski and Krawczyk 2024; Vintskevich et al. 2023). These applications underscore the adaptability and potential scalability of machine learning approaches to the nuanced analysis of quantum states. The challenge extends to not only identifying entangled states but also understanding the complex web of quantum correlations within multipartite systems such as quantum non-locality. Here, machine learning methodologies, such as those leveraging Bell's inequalities (Schatzki et al. 2021), and the development of separability-entanglement classifiers (Sirui et al. 2018), represent significant advancements in the field. They offer a more resource-efficient alternative to comprehensive quantum state analysis methods, like quantum state tomography, by processing partial information about quantum states.

✉ Giuseppe Sergioli
giuseppe.sergioli@gmail.com

1 Università degli Studi di Cagliari, Via Is Mirrions, 09123 Cagliari, Italy

2 Technische Universität München. Institute for Advanced Study (IAS), Lichtenbergstraße 2a, 85748 Garching b. München, Germany

3 Technische Universität München. Department of Science, Technology and Society, Arcisstr. 21, 80333 München, Germany

4 European Organization for Nuclear Research (CERN), Espl. des Particules 1, 1211 Meyrin, Switzerland

5 Technische Universität München. Department of Engineering and Design, Boltzmannstr. 15, 85748 Garching b. München, Germany

Published online: 25 November 2025

Furthermore, the integration of machine learning with quantum entanglement theory has facilitated the development of novel strategies for entanglement classification and quantification. For instance, the use of neural network quantum states within a restricted Boltzmann machine architecture (Harney et al. 2020) highlights the innovative strategies being developed to address the complexities of quantum correlations. These efforts not only enhance our understanding of quantum states but also pave the way for practical applications in quantum computing and information processing.

Another direction of research considers the employing of quantum-specific datasets (Ma and Yung 2018) to train QML models. This pivot from classical to quantum datasets turned out to be important, given the complex and high-dimensional nature of quantum information, which classical datasets fail to fully represent. By focusing on quantum states characterized by various degrees of multipartite entanglement, these studies have begun to uncover the advantages that QML models may offer, particularly in tasks like the supervised learning classification of quantum states. The comparison of models trained on different types of data (Lin et al. 2023) sheds light on the trade-offs and considerations necessary for optimizing machine learning models for quantum applications. The exploration of hybrid quantum-classical machine learning frameworks further exemplifies the field's forward momentum, aiming to improve model performance through the generation of informative correlation data and showcasing a noise-resistant capability crucial for real-world applications. In sum, the state of the art in leveraging quantum machine learning for the classification and quantification of entanglement reflects a dynamic and rapidly evolving field. It embodies the fruitful intersection of quantum physics and artificial intelligence, offering new insights into quantum entanglement and opening up novel avenues for quantum computing.

In this work, we propose yet another new approach inspired by quantum state discrimination (Giuntini et al. 2023a, b). Specifically, after briefly summarizing the workings of the classifier based on the Pretty Good Measurement (PGM) (Hausladen and K. Wootters 1994; Mochon 2006), we describe the setup of the experiment and introduce some notational conventions that allow to distinguish among factorized states, separable states, and entangled states. In particular, reference will also be made to various types of factorizability, with the aim of achieving a refined classification capable of capturing these differences.

The PGM Classifier will be the key instrument in our research to distinguish not only among factorized, separable, and entangled states but also to identify those states that exhibit non-locality by violating the Mermin and Svetlichny inequalities (Svetlichny (1987); Mermin 1990). We will create a complete set of states that serve both entanglement and non-locality. In this way, we ensure that our classification is founded on the idea of being able to identify genuine or non-genuine quantum correlations that go beyond classical locality. Applying Mermin and Svetlichny inequalities not only deepens our understanding of quantum states but also establishes a robust framework for analyzing and identifying unique peculiarities in quantum correlations within multi-qubit systems. In summary, we will utilize the PGM Classifier on a dataset of quantum states to compare its results with those obtained from traditional classifiers, emphasizing the effectiveness and advantages of our quantum approach. Further, a detailed account of the general benefits of the PGM-based classification strategy is summarized in the discussion section.

The article is organized as follows. In Section 2, we will present a brief introduction to the PGM Classifier. In Section 3, we will explore the theoretical framework, giving a detailed description of the different quantum correlations that will be considered. In Section 4, we will address the notation and methodology for classifying factorized, separable, and entangled states. Following the same thought, Section 5 will be devoted to classifying states based on the non-locality violation (specifically using the Mermin and Svetlichny inequalities as a criterion). Finally, in Section 6, we draw our conclusions.

2 The PGM Classifier

Quantum information theory, and particularly the theory of quantum state discrimination, have inspired the development of some quantum-inspired classification algorithms[1] that, in comparison to other classical classifiers, have shown very promising performance in terms of accuracy (Sergioli 2020; Sergioli et al. 2019). Drawing inspiration from PGM (Hausladen and K. Wootters 1994; Mochon 2006), it has been possible to define a quantum-inspired multi-class classifier that exhibits good performance both in terms of accuracy and computational complexity (running time). The definition of the PGM Classifier has been detailed in some previous works, where the performance of the quantum-inspired classifier was also compared to other standard classifiers. To make the article self-contained, in this section, we briefly summarize the description of the PGM Classifier (Giuntini et al. 2023a, b).

Let us consider a training dataset represented as a collection of patterns:

$$\mathcal{S}_{tr} := \{(\mathrm{x}_1, y_1), \ldots, (\mathrm{x}_m, y_m)\}\,, \tag{1}$$

where $\mathrm{x}_j \in \mathbb{C}^d$ denotes a d-dimensional feature of the complex Hilbert space $\mathbb{C}^d$ and $y_j \in \mathcal{Y} := \{0, \ldots, \ell\}$ for all

[1] Conversely, quantum algorithms are designed to run on quantum devices.

$j \in \{1, \ldots, m\}$ signifies the associated class label. For a given class label $i \in \mathcal{Y}$, the subset $\mathcal{S}_{tr}^i$ comprising all feature vectors affiliated with class label i is defined as:

$$\mathcal{S}_{tr}^i = \{\mathrm{x}_j \in \mathcal{S}_{tr} : y_j = i\}\,. \tag{2}$$

The cardinality of $\mathcal{S}_{tr}^i$ is represented as $|\mathcal{S}_{tr}^i|$, leading to the relationship $\sum_{i=1}^{\ell} |\mathcal{S}_{tr}^i| = m$. Within this framework, the objective of supervised classification is to deduce a function from the training dataset $\mathcal{S}_{tr}$, employing a classifier to optimize the precision of class label assignment for an unseen (i.e., unlabeled) feature vector x. A multi-class classifier is devised by identifying a mapping that correlates any feature vector x with a sequence of ℓ-numbers within the unit real interval $[0, 1] \subset \mathbb{R}$, expressed as $f : \mathbb{C}^d \to [0, 1]^\ell$. The i-th element of $f(\mathrm{x})$, denoted by $f(\mathrm{x})_i$, represents the classification score for assigning label i to the data. In case the range of $f(\mathrm{x})$ is $[0, 1] \subset \mathbb{R}$ and $\sum_{i=1}^{\ell} f(\mathrm{x})_i = 1$, we may interpret the score as the probability of x belonging to class i.

The classifier, defined by f (or the f-classifier), is a mapping $Cl_f : \mathbb{C}^d \to \mathcal{Y}$, which assigns to any feature vector $\mathrm{x} \in \mathbb{C}^d$ the class label correlated with the highest score in $f(\mathrm{x})_i$, for $1 \le i \le \ell$. That is, $Cl_f(\mathrm{x}) = \underset{i}{\mathrm{argmax}}\{f(\mathrm{x})_i : 1 \le i \le \ell\}$.

Given the possibility of multiple maximum values in $f(\mathrm{x})$, by convention, we refine the definition of our multi-class classifier as:

$$Cl_f(\mathrm{x}) := \min\left\{i \in \mathcal{Y} : f(\mathrm{x})_i = \max_k\{f(\mathrm{x})_k, 1 \le k \le \ell\}\right\}. \tag{3}$$

The construction of the PGM Classifier necessitates evaluating $f(\mathrm{x})$ through the prism of quantum theoretical principles. To bridge classical and quantum realms, each feature vector x is represented as a quantum state via a density operator ρ_{x} (we will say a *quantum pattern*). This process of quantum feature mapping transforms classical information encapsulated in x into a quantum state, employing methodologies proposed in prior research (Sergioli et al. 2017; Santucci and Sergioli 2018; Sergioli 2020).

Let us define a *quantum pattern* as a pair $(\rho_{\mathrm{x}_j}, y_j)$. The quantum training dataset, comprising all such quantum patterns, is represented as $\mathcal{S}_{Qtr} = \{(\rho_{\mathrm{x}_1}, y_1), \ldots, (\rho_{\mathrm{x}_m}, y_m)\}$. For class label $i \in \mathcal{Y}$, $\mathcal{S}_{Qtr}^i$ denotes the set of all quantum states ρ_{x_j} corresponding to feature vectors in class i, articulated as $\mathcal{S}_{Qtr}^i = \{\rho_{\mathrm{x}_j} : \mathrm{x}_j \in \mathcal{S}_{tr}^i\}$.

The *quantum centroid* for class i is defined as the uniformly weighted convex combination of all density operators within the class, described by

$$\rho_{(i)} = \frac{1}{|\mathcal{S}_{Qtr}^i|} \sum_{\mathrm{x}_j \in \mathcal{S}_{tr}^i} \rho_{\mathrm{x}_j}\,, \tag{4}$$

where $|\mathcal{S}_{Qtr}^i|$ equals the cardinality of $\mathcal{S}_{tr}^i$. Thus, each of the ℓ class labels corresponds to a unique quantum centroid. To extend the concept of a quantum centroid, we encode the classical vectors of the dataset into multiple copies of a density operator $\rho_{\mathrm{x}}^{\otimes n} := \underbrace{\rho_{\mathrm{x}} \otimes \cdots \otimes \rho_{\mathrm{x}}}_{n\ times}$, where $\otimes$ is tensor product. In this case, the generalized quantum centroid is expressed by:

$$\rho_{(i)}^{(n)} = \frac{1}{|\mathcal{S}_{Qtr}^i|} \sum_{\mathrm{x}_j \in \mathcal{S}_{tr}^i} \rho_{\mathrm{x}_j}^{\otimes n}\,. \tag{5}$$

This approach, informed by kernel theory (Mengoni and Di Piero 2019), posits that classifying a non-linear dataset is feasible via a separating hyperplane in a higher-dimensional space, though the optimal number of copies (tensor products) requires careful selection to balance computational time against classification accuracy.

After the encoding and the calculation of the quantum centroids for each class, we apply quantum state discrimination (QSD) (Barnett and Croke 2009) techniques to obtain the score function f. QSD, which assesses the distinguishability of quantum states, provides a foundation for differentiating quantum centroids in multi-class classification scenarios.

Let us consider a collection of quantum states accompanied by their corresponding prior probabilities, $R = \{(\rho_1, \mathrm{p}_1), \ldots, (\rho_\ell, \mathrm{p}_\ell)\}$. The mean state of R is described by: $\sigma = \sum_{i=1}^{\ell} \mathrm{p}_i \rho_i$, and for each i, where $1 \le i \le \ell$, we define a set of operators $\{E_i\}$ as follows:

$$E_i = (\sigma^\dagger)^{1/2} \mathrm{p}_i \rho_i (\sigma^\dagger)^{1/2}\,,$$

with $\sigma^\dagger$ representing the pseudoinverse (or Moore-Penrose inverse) of σ. For every i ranging from 1 to ℓ, the operators are determined by

$$F_i = E_i + \frac{1}{\ell} P_{\ker(\sigma)}\,, \tag{6}$$

where $P_{\ker(\sigma)}$ is the projection onto the subspace defined by the kernel of σ.

We employ the PGM formalism for minimum-error discrimination, constructing the set of measurement operators $\{F_i\}$ for the ensemble $R = \{(p_i, \rho_{(i)}^{(n)}), \ldots, (p_\ell, \rho_{(\ell)}^{(n)})\}$,

where p_i represents the proportion of the training dataset within class i.

The PGM Classifier for an unknown quantum datum $\rho_{\mathrm{x}}^{(n)}$ is thus defined as

$$Cl_f(\mathrm{x}) := \min_i \left\{ i \in \mathcal{Y} : p_i \mathrm{Tr}\left(F_i \rho_{\mathrm{x}}^{(n)}\right) = \max_k \{p_k \mathrm{Tr}\left(F_k \rho_{\mathrm{x}}^{(n)}\right), 1 \leq k \leq \ell\} \right\}. \tag{7}$$

This framework elucidates the application of QSD to the classification challenge, highlighting that enhanced QSD success probabilities between quantum centroids correlate with improved classification accuracy (Giuntini et al. 2023a, b). Consequently, empirical and theoretical explorations suggest that increasing the number of tensor copies (n) can elevate the classifier's performance.

Let us remark that our classification process also incorporates a rescaling pre-processing step, which consists in multiplying each component of every vector x in the original dataset by a positive real parameter α. Naturally, due to the encoding procedure, the quantum centroid is not invariant under rescaling. Therefore, performing a grid search optimization over the parameter α can lead to a further improvement in classification performance in terms of accuracy, as demonstrated in Sergioli et al. (2017).

The workflow of the PGM classifier is depicted in Fig. 1.

3 Theoretical background

In the field of quantum information, identifying quantum correlations is of paramount importance. This involves distinguishing between different types of quantum states and their properties. Specifically, it is crucial to differentiate between: (i) product (or factorized) states, separable states, and entangled states, and (ii) states that either violate or do not violate non-locality. Non-locality is typically determined through the violation of the Mermin or Svetlichny inequalities.

This section is devoted to fixing some notation, while the setup of the experiment is described in the next section.

In the realm of quantum mechanics, the classification of pure states into factorized, separable, and entangled states provides crucial insights into the nature of quantum correlations.

1. **Fully separable** (or **Factorized) states:** A pure state of a n-dimensional composite quantum system is said to be *fully separable* or *factorized* iff it can be written by using $(n-1)$ tensor products of a number n of 1-qubit systems. Formally, the n-dimensional qubit system $|\psi\rangle$ will be written as $|\psi\rangle = |\psi\rangle_1 \otimes \cdots \otimes |\psi\rangle_n$, where $|\psi\rangle_i$ are qubits.
2. **Partially separable** (for the sake of simplicity we refer to **Separable) states:** A pure state of a n-dimensional

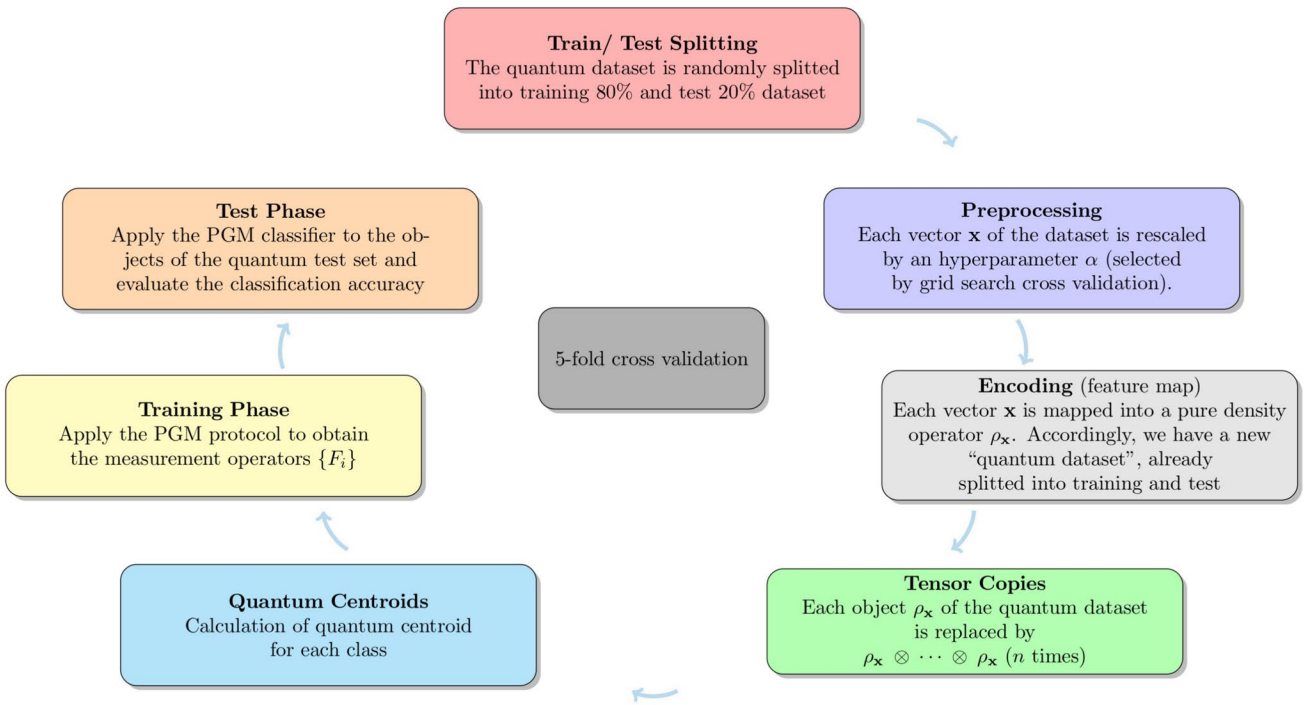

Fig. 1 The workflow and train/test procedure of the PGM classifier

composite quantum system (with $n > 2$) is said to be *partially separable* or, simply, *separable* iff it can be written as m tensor products ($1 \leq m < n-1$) of $(m+1)$ subsystems such that: *i*) each subsystem i has dimension k_i (such that $1 \leq k_i < n$ and $\sum_i k_i = n$) and *ii*) at least one subsystem i is not a factorized state. As an example, a pure 3-dimensional ($n = 3$) composite quantum system $|\psi\rangle$ can be written as one tensor product ($m = 1$) of two subsystems $|\psi\rangle = |\psi\rangle_1 \otimes |\psi\rangle_2$ such that *i*) one subsystem has dimension 1 and the other one has dimension 2 (indeed $1 \leq k < 3$) and *ii*) one of the subsystem $|\psi\rangle_1$ or $|\psi\rangle_2$ will be not factorized.

3. **Entangled States:** A pure state of a n-dimensional composite quantum system is said to be *entangled* iff it is neither a factorized nor a separable state. In other words, it can not be written as a tensor product of any subsystem.

The distinction between these classes of states underpins many quantum phenomena and it is crucial quantum information theory, including applications in quantum computing, quantum cryptography, and quantum teleportation. In the following section, we provide a detailed specification of the various classes of product, separable, and entangled states to which we apply our quantum-inspired classifier, taking into account the differing numbers of qubits considered in each case.

We also examine multipartite non-locality by analyzing quantum correlations through entanglement and separability measures. For the case of two qubits, we will use the Clauser-Horne-Shimony-Holt (CHSH) inequality (Clauser et al. 1969), a well-known Bell-type inequality. For three qubits, we will use the Mermin (Mermin 1990) and Svetlichny inequalities (Svetlichny (1987); Seevinck and Svetlichny 2002).

Svetlichny's work introduces an inequality that is violated by quantum mechanics and is capable of detecting three-qubit correlations that cannot be reduced to mixtures of locally related two-qubit correlations with the third. This inequality distinguishes between the non-separability of three qubits and that of two-qubit, akin to Bell-type inequalities. We derive these inequalities under the limited entanglement hypothesis, employing alternative dichotomic observables for each subsystem to identify specific maximum values in Eq. 10.

In our framework, a violation of Bell-type inequalities signals that the observed quantum correlations cannot be accounted for by local realistic models. This provides a principled basis for our classification scheme, allowing us to distinguish genuine quantum correlations from those compatible with classical locality

Table 1 Possible configurations for each case of a 2-qubit system

Factorized	Entangled
ab	*AA*

Table 2 Possible configurations for each case of a 3-qubit system

Factorized	Separable	Entangled
abc	*AAb*	*AAA*
	AbA	
	bAA	

4 Classifying Factorized, Separable, and Entangled quantum states

The learning phase of the algorithm was carried out by leveraging the structure of the vectors (pure states) used to describe the multi-qubit systems considered in our analysis. Consequently, we generated datasets of random vectors representing each case we describe, now depending on the varying number of qubits considered at each instance. In this section, we describe in detail the different classes of factorized, separable, and entangled states to which we applied our quantum-inspired classifier, taking into account the differing numbers of qubits considered in each case.

First, let us consider the simpler 2-qubit case. In this scenario, only two possible cases exist: i) the two qubits represent a factorized state; ii) the two qubits represent an entangled state.

For simplicity, we employ the following pictorial representation, where lowercase letters denote uncorrelated qubits, while uppercase letters denote correlated qubits. Thus, case I) is represented as $a\ b$ and case II) as $A\ A$ (in the latter case, the same uppercase letter is used to indicate that the two qubits form an entangled state).

We indicate this case as in Table 1.

By referring to the notation above, trivially, we will not have separable states for $n = 1$ because $m = 0$.

Now let us consider the three qubits case (see Table 2). In this case, $m = 1$; hence, we have separable states (as discussed above). Three sub-cases arise in this scenario:

1. None of the three qubits are related to each other: abc.
2. All three qubits are related to each other: AAA.
3. Two qubits are related to each other, but not with the remaining qubit: AAb.

To avoid bias in the learning process, during dataset generation, all possible configurations in which case III. can occur must be considered. Particularly, case III. can occur in three different configurations:

3.1 AAb.
3.2 AbA.

Table 3 Possible configurations for each case of a 4-qubit system

Factorized	Separable type I	Separable type II	Separable type III	Entangled
abcd	*AAAb*	*AABB*	*AAbc*	*AAAA*
	AAbA	*ABAB*	*AbAc*	
	AbAA	*ABBA*	*AbcA*	
	bAAA		*bAcA*	
			bcAA	
			bAAc	

3.3 bAA.

Configurations 3.2 and 3.3 can be readily obtained from 3.1 through appropriate applications of Swap operators.

For the case of four qubits, we will distinguish not only between factorized, separable, and entangled states but will also provide a detailed classification among different types of separable states, as shown in Table 3. Specifically, for $n = 4$, *m* can assume the values 1 or 2, and *k* can be 1, 2, or 3. This allows us to represent all different types of separable states. In details, for $m = 1$, $k_1 = 3$, and $k_2 = 1$, we have *Separable type I*; for $m = 1$, $k_1 = 2$, and $k_2 = 2$, we have *Separable type II* (where we have two separated two-qubit entangled systems, indicated with different capital letters AA and BB, respectively); finally, for $m = 2$, $k_1 = 2$, $k_2 = 1$, and $k_3 = 1$, we have *Separable type III*. As will be shown, the classification proves to be particularly challenging, especially for separable states. Therefore, in the next section, we will first present the results of the classification among factorized, separable, and entangled states, but we will also analyze the outcomes of a finer classification among the various types of separable states.

As can be seen from Table 3, for instance, type I separable states are composed of 4 qubits, among which 3 are entangled with each other, while the remaining one is uncorrelated. All other cases can be read and interpreted in a similar manner. Once again, all the cases included in the same type of separability can be obtained by suitable permutations, i.e., trivial application of Swap gates.

Also, the five qubits case can be depicted in a very similar way (see Table 4). In this scenario, different cases are given. The value of *m* can be equal to 1 (*Separable type* I - with $k_1 = 4$ and $k_2 = 1$ -and *Separable type* V- with $k_1 = 2$ and $k_2 = 3$). The value of *m* can be equal to 2 (*Separable type* II - with $k_1 = 3$, $k_2 = 1$ and $k_3 = 1$ -and *Separable type* IV- with $k_1 = 2$, $k_2 = 2$ and $k_3 = 1$). Finally, *m* can be equal to 3 (*Separable type* III, with $k_1 = 2$, $k_2 = 1$, $k_3 = 1$ and $k_4 = 1$).

4.1 Experimental results

We now describe in detail the experimental setup employed in our study. For each scenario discussed in the previous section–namely, systems composed of 2, 3, 4, or 5 qubits–we generated random quantum states as input data. The classification experiments were repeated for multiple dataset sizes: 10,000; 15,000; 20,000; 25,000; and 30,000 samples. In all cases, the dataset was partitioned into an 80% training set and a 20% test set. Model hyperparameters were optimized using a grid search with 5-fold cross-validation performed exclusively on the training set. As outlined in Section 2, two key hyperparameters influence the performance of the PGM classifier. The first is the number of copies of the quantum state available for classification. Prior studies (Sergioli et al. 2019) have shown that increasing the number of copies improves the probability to increase the classification accuracy. However, due to memory limitations of the computational devices used, we restricted the number of copies to a maximum of two across all experiments. The second relevant hyperparameter is the rescaling factor α, which, as discussed at the end of Section 2, can also affect the performance of the PGM

Table 4 Possible configurations for each case of a 5-qubit system

Factorized	Sep. t. I	Sep. t. II	Sep. t. III	Sep. t. IV	Sep. t. V	Entangled
abcde	*AAAAb*	*AAAbc*	*AAbcd*	*AABBc*	*AABBB*	*AAAAA*
	AAAbA	*AAbAc*	*AbAcd*	*ABABc*	*ABAAB*	
	AAbAA	*AAbcA*	*bAAcd*	*ABBAc*	*ABABA*	
	AbAAA	*AbAAc*	*AbcAd*	*AABcB*	*ABBAA*	
	bAAAA	*AbAcA*	*bAcAd*	*ABAcB*	*AABBA*	
		AbcAA	*bcAAd*	*ABBcA*	*AAABB*	
		bAAAc	*AbcdA*	*AAcBB*	*BAAAB*	
		bAAcA	*bAcdA*	*ABcAB*		
		bAcAA	*bcAdA*	*ABcBA*		
		bcAAA	*bcdAA*	*AcABB*		
				AcBAB		
				AcBBA		
				cAABB		
				cABAB		
				cABBA		

classifier. To simplify the computational load, we performed a grid search over two values of the rescaling parameter, specifically $\alpha \in \{0.5, 1\}$. This concludes the description of the PGM classifier setup. In order to facilitate a comparison with standard machine learning classifiers from the `scikit-learn` library, we applied a *floating* transformation to the original quantum states. Formally, a pure quantum state $|\psi\rangle$ of a n-dimensional composite quantum system can be described by a 2^n- dimensional vector of, generally, complex numbers x_i. It can be easily transformed into a 2^{n+1}-dimensional real vector obtained by decomposing each complex number x_i into its real and imaginary parts. The hyperparameter search space - as included in `scikit-learn` - is summarized in Table 12 (see Appendix at the end of the paper) and the best configurations for all the classifiers compared along the experiments are summarized in Table 13 and Table 14 for the two classification tasks: (i) Factorized/Separable/Entangled (as described in the previous section) and (ii) Local/Non-Local (that will be described in Section 5).

Here, we present the numerical results of our experiments, designed to compare the performance of the PGM Classifier with **two tensor copies** against various standard classifiers, in distinguishing between Factorized, Separable and Entangled states. Experiments were conducted on systems comprising 2, 3, 4, and 5 qubits. The average balanced accuracy of each classifier across all test sets, as shown in Tables 5, 6, 7, and 8, highlights the strong performance of the PGM Classifier compared to the other classifiers evaluated. Individual heatmaps illustrating the balanced accuracy achieved by the best-trained model of each classifier on each dataset are presented in Figs. 2, 5, 8, and 11.

Additionally, the biclustering heatmaps shown in Figs. 3, 6, 9 and 12, comparing classifiers for each dataset, indicate that the average winning ratio of the PGM Classifier exceeds that of most of all other classifiers. Finally, the confusion matrices shown in Figs. 4, 7, 10, and 13 provide a detailed analysis of the classifier's performance. These matrices illustrate the distribution of true positive, false positive, true negative, and false negative predictions for each class, highlighting the ability of the classifiers to correctly identify factorized, separable, and entangled quantum states. They complement the balanced accuracy metric by offering deeper insight into how each classifier distinguishes between these different states (Figs. 5, 6, 7, 8, 9, 10, 11, 12, 13).

4.1.1 2-Qubit case

Table 5 Average Balanced Accuracy for each classifier

Classifier	Average Balanced Accuracy
MLP	0.935 ± 0.026
Support Vector Machine (SVM)	0.910 ± 0.011
PGM	0.867 ± 0.005
Nearest Neighbors (NN)	0.841 ± 0.017
Helstrom Bound Classifier	0.735 ± 0.005
Random Forest (RF)	0.743 ± 0.011
Extra Tree (ET)	0.726 ± 0.014
Nearest Centroid (NC)	0.505 ± 0.002
Quadratic Discriminant Analysis (QDA)	0.504 ± 0.003
Dummy Classifier	0.504 ± 0.003
Linear Discriminant Analysis (LDA)	0.503 ± 0.002
Bernoulli Naive Bayes (B-NB)	0.502 ± 0.005
Gaussian Naive Bayes (G-NB)	0.500 ± 0.011

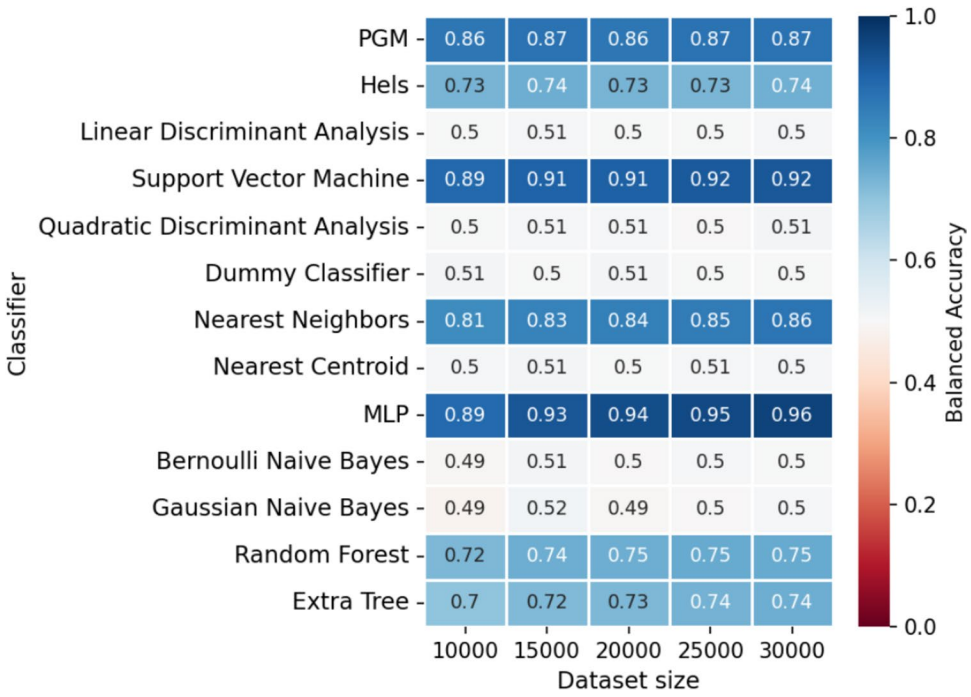

Fig. 2 Heatmap showing the Balanced Accuracy for different classifiers in the 2-qubit case, as a function of the number of samples

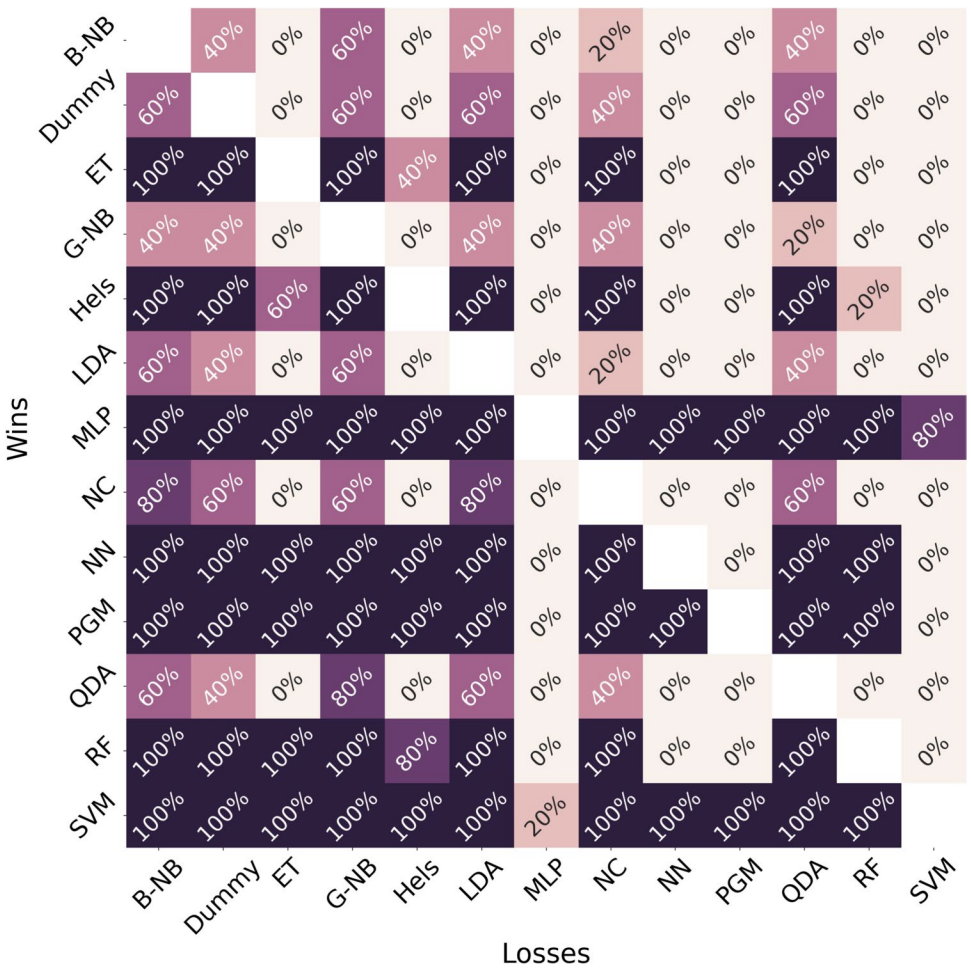

Fig. 3 Biclustering map, when a classifier A outperforms a classifier B according to the Balanced Accuracy. A lighter color indicates a lower percentage

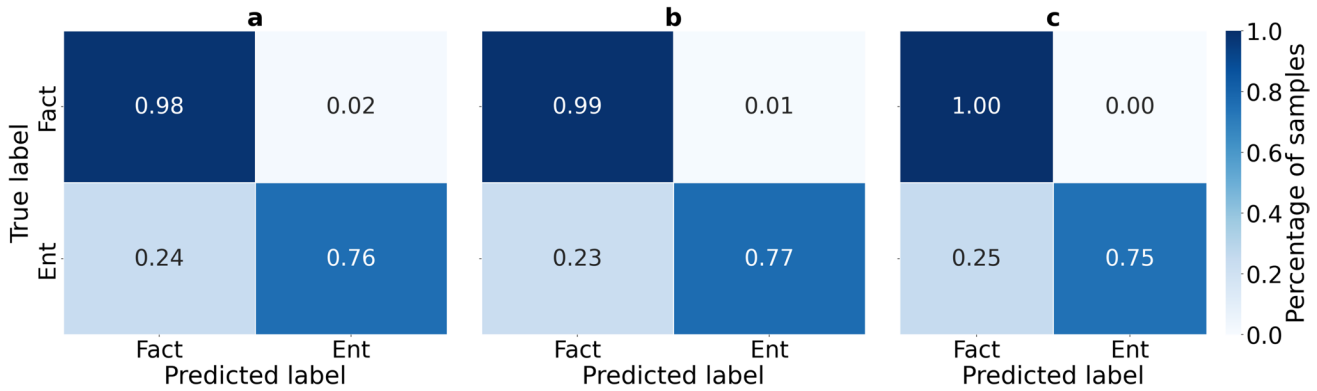

Fig. 4 Confusion matrix for a 2-qubit system with two classes for: (a) 10,000, (b) 20,000, and (c) 30,000 samples

4.1.2 3-Qubit case

Table 6 Average Balanced Accuracy for each classifier

Classifier	Average Balanced Accuracy
Support Vector Machine (SVM)	0.744 ± 0.041
PGM	0.694 ± 0.007
Nearest Neighbors (NN)	0.623 ± 0.020
Random Forest (RF)	0.631 ± 0.014
Extra Tree (ET)	0.589 ± 0.014
MLP	0.561 ± 0.036
Nearest Centroid (NC)	0.340 ± 0.003
Quadratic Discriminant Analysis (QDA)	0.344 ± 0.005
Dummy Classifier	0.336 ± 0.002
Gaussian Naive Bayes (G-NB)	0.336 ± 0.009
Bernoulli Naive Bayes (B-NB)	0.334 ± 0.004
Linear Discriminant Analysis (LDA)	0.334 ± 0.004

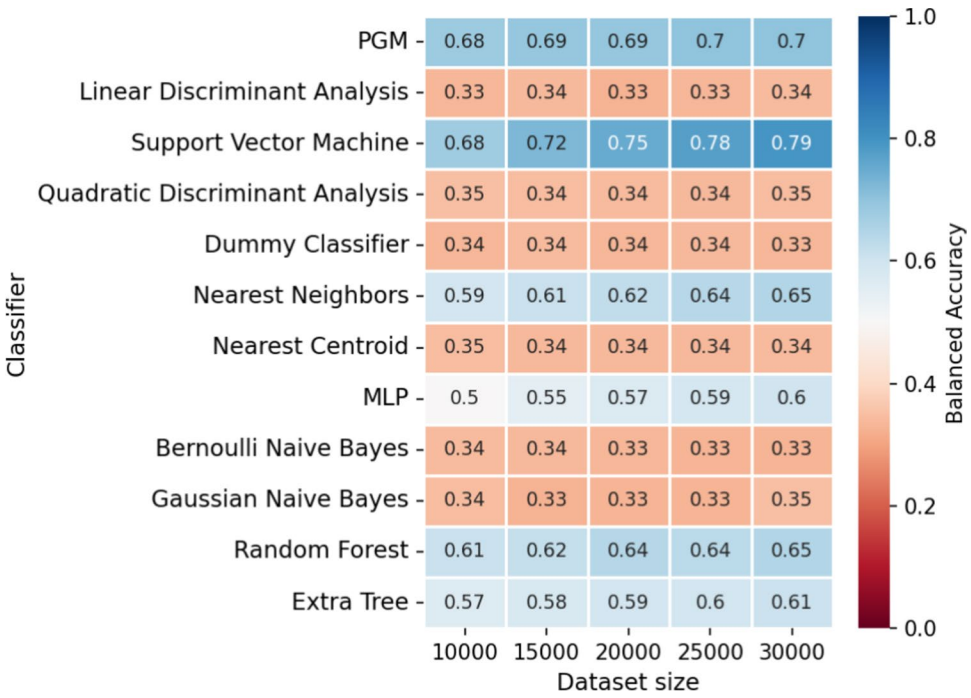

Fig. 5 Heatmap showing the Balanced Accuracy for different classifiers in the 3-qubit case, as a function of the number of samples

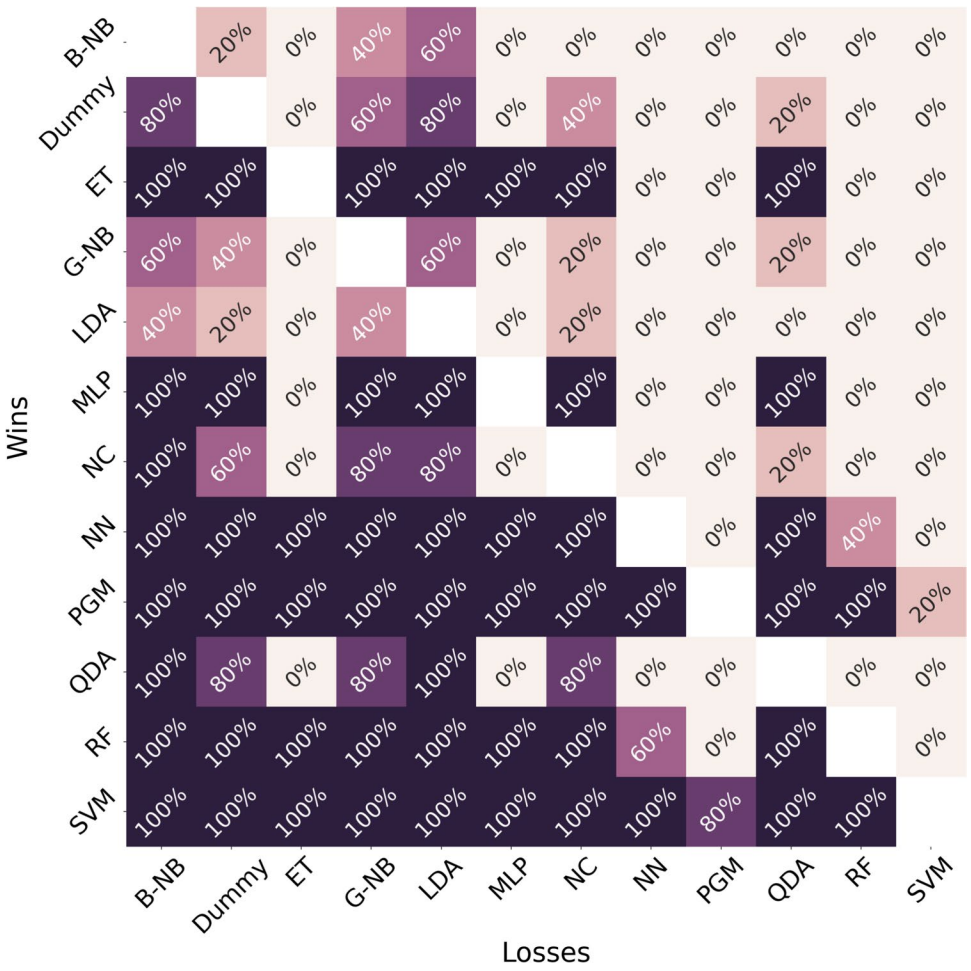

Fig. 6 Biclustering map, when a classifier A outperforms a classifier B according to the Balanced Accuracy. A lighter color indicates a lower percentage

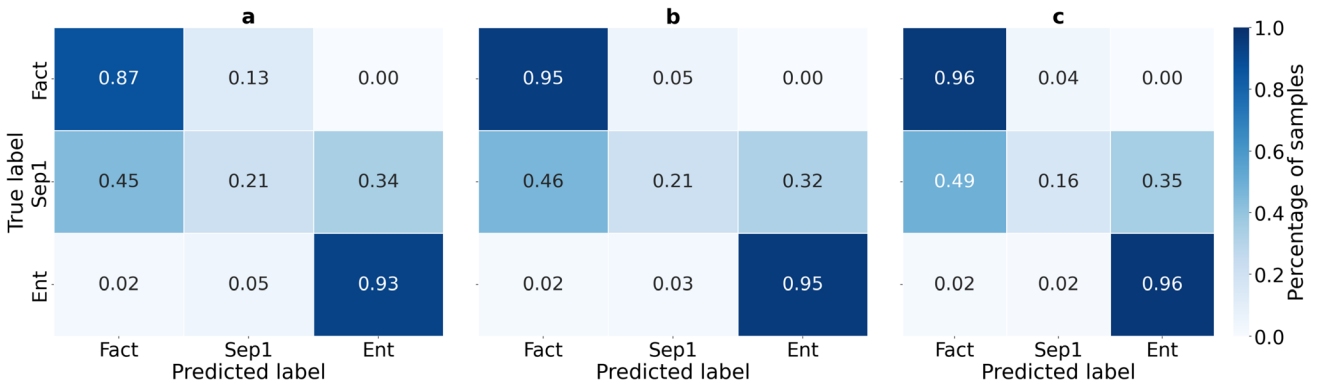

Fig. 7 Confusion matrix for a 3-qubit system with three classes for: (a) 10,000, (b) 20,000, and (c) 30,000 samples

4.1.3 4-Qubit case

Table 7 Average Balanced Accuracy for each classifier

Classifier	Average Balanced Accuracy
PGM	0.470 ± 0.015
Random Forest (RF)	0.418 ± 0.015
Extra Tree (ET)	0.385 ± 0.007
Support Vector Machine (SVM)	0.354 ± 0.043
Nearest Neighbors (NN)	0.354 ± 0.014
MLP	0.244 ± 0.023
Quadratic Discriminant Analysis (QDA)	0.210 ± 0.003
Gaussian Naive Bayes (G-NB)	0.209 ± 0.007
Nearest Centroid (NC)	0.204 ± 0.004
Dummy Classifier	0.202 ± 0.001
Bernoulli Naive Bayes (B-NB)	0.201 ± 0.003
Linear Discriminant Analysis (LDA)	0.198 ± 0.002

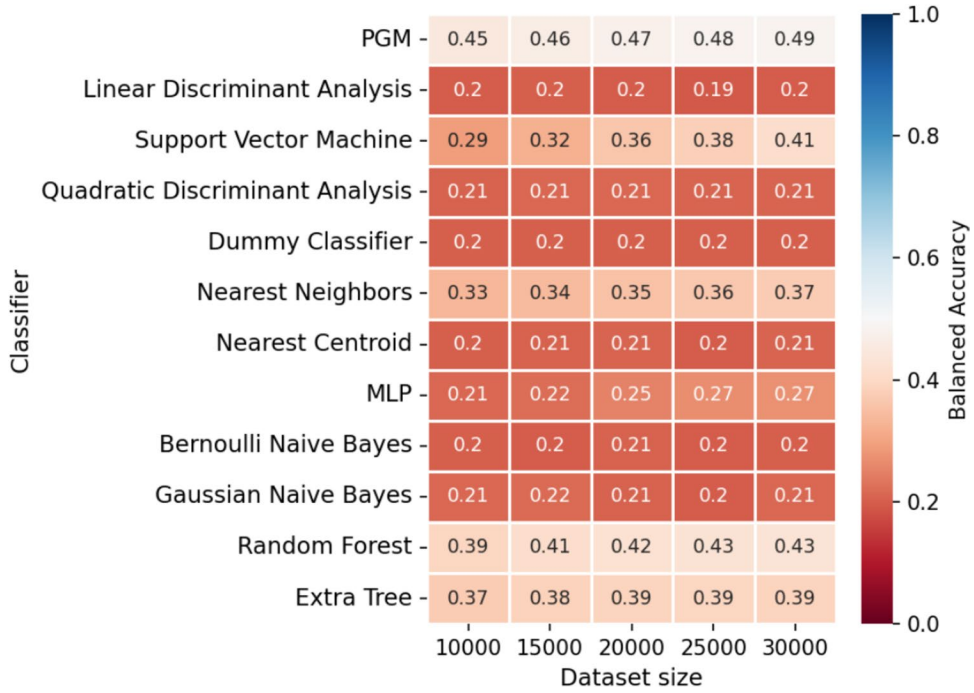

Fig. 8 Heatmap showing the Balanced Accuracy for different classifiers in the 4-qubit case, as a function of the number of samples

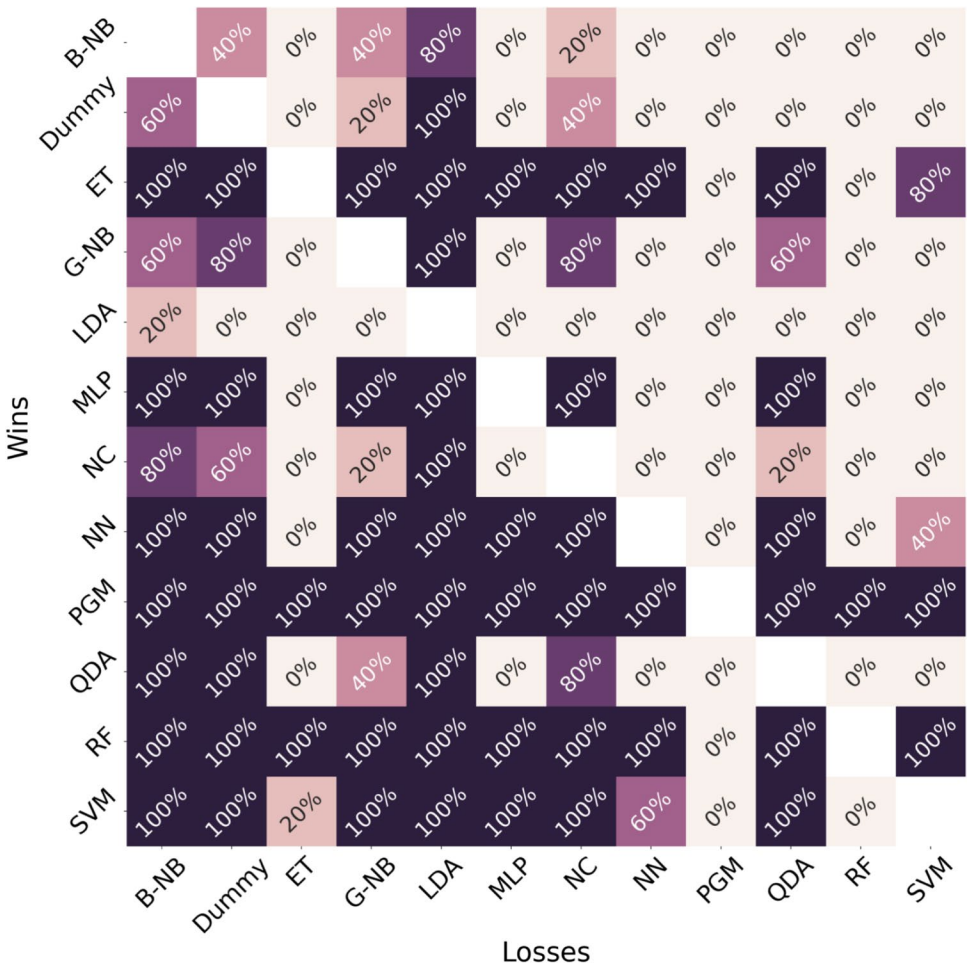

Fig. 9 Biclustering map, when a classifier A outperforms a classifier B according to the Balanced Accuracy. -A lighter color indicates a lower percentage

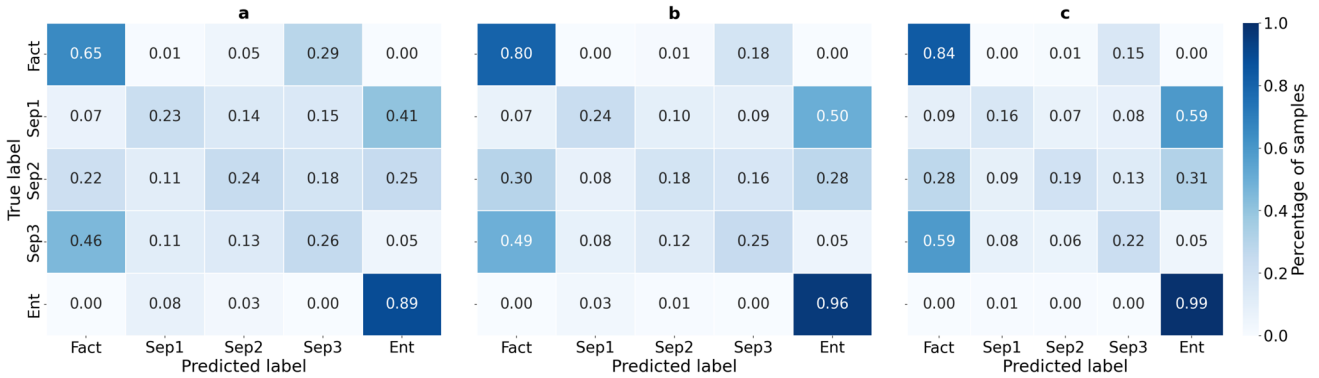

Fig. 10 Confusion matrix for a 4-qubit system with five classes for: (a) 10,000, (b) 20,000, and (c) 30,000 samples

4.1.4 5-Qubit case

Table 8 Average Balanced Accuracy for each classifier

Classifier	Average Balanced Accuracy
PGM	0.377 ± 0.023
Random Forest (RF)	0.319 ± 0.008
Extra Tree (ET)	0.296 ± 0.003
Nearest Neighbors (NN)	0.244 ± 0.006
Support Vector Machine (SVM)	0.194 ± 0.012
Quadratic Discriminant Analysis (QDA)	0.155 ± 0.002
MLP	0.153 ± 0.005
Gaussian Naive Bayes (G-NB)	0.152 ± 0.005
Nearest Centroid (NC)	0.149 ± 0.003
Dummy Classifier	0.147 ± 0.005
Linear Discriminant Analysis (LDA)	0.146 ± 0.004
Bernoulli Naive Bayes (B-NB)	0.143 ± 0.003

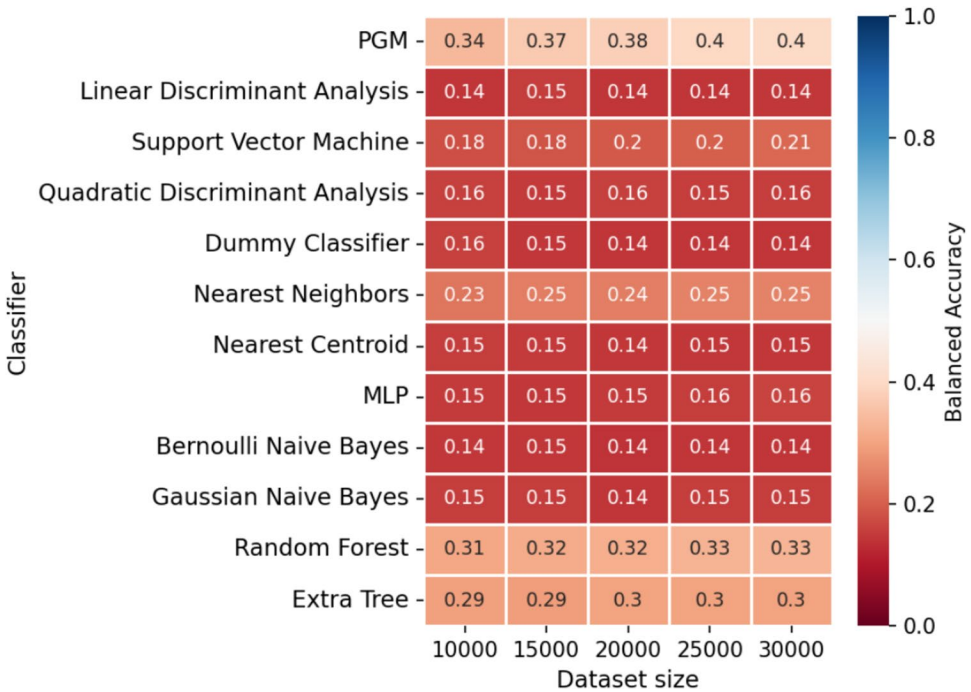

Fig. 11 Heatmap showing the Balanced Accuracy for different classifiers in the 5-qubit case, as a function of the number of samples

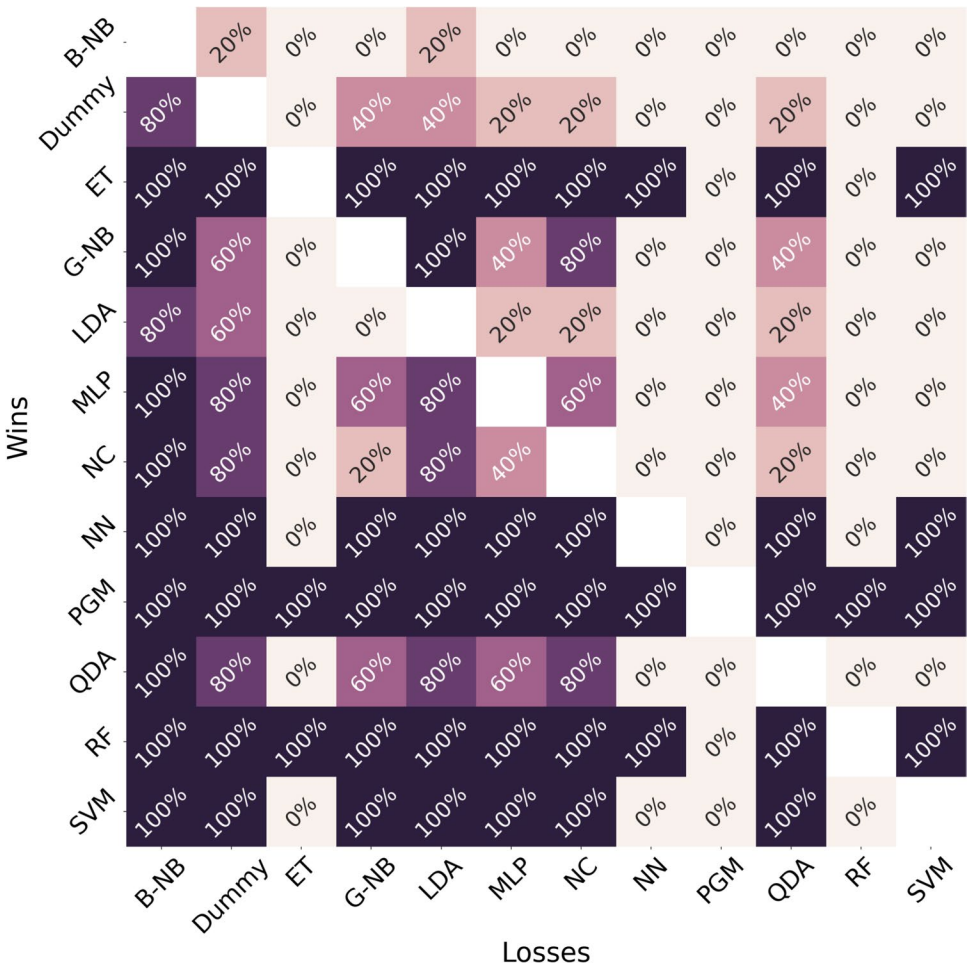

Fig. 12 Biclustering map, when a classifier A outperforms a classifier B according to the Balanced Accuracy. A lighter color indicates a lower percentage

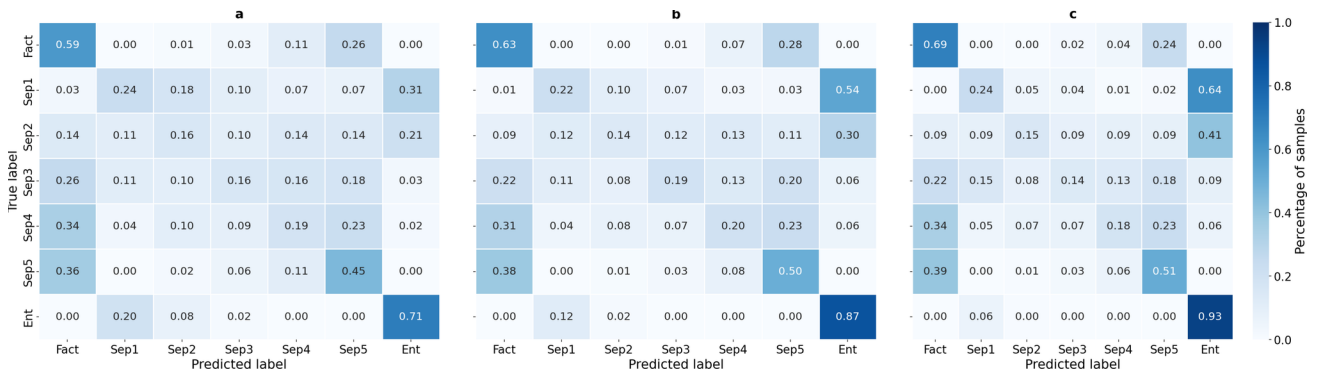

Fig. 13 Confusion matrix for a 5-qubit system with seven classes for: (a) 10,000, (b) 20,000 and (c) 30,000 samples

5 Classification of non-local quantum states

Non-locality is a cornerstone of quantum mechanics that challenges classical intuitions regarding the interactions, mutual influence, and correlations between particles. A key result demonstrating non-locality is the violation of Bell's inequalities, which distinguishes the predictions of quantum mechanics from those of theories based on local hidden variables.

Bell's theorem, first formulated in 1964 (Bell 1964; Brunner et al. 2014), proves that no local hidden variable theory can fully replicate the predictions of quantum mechanics. Specifically, if quantum mechanical predictions hold, measurements performed on entangled particles can exhibit correlations that are intrinsically non-local.

The strength and nature of these non-local correlations are typically quantified using Bell-type inequalities. For two-qubit systems, the most well-known is the CHSH inequality (Clauser et al. 1969). For systems involving more qubits, the Mermin inequalities (Mermin 1990) and the Svetlichny inequalities (Svetlichny (1987)) provide extensions that further characterize multi-qubit non-local correlations.

This formalism not only highlights the departure of quantum mechanics from local realism but also serves as a critical tool for understanding and harnessing entanglement in quantum technologies, including quantum computation and quantum machine learning. In this section, we focus on the classification of states based on the violation of these inequalities.

We begin by considering a physical system composed of two subsystems. In this framework, we define two dichotomic observables, A_1 and A_1', acting on the first subsystem, and two dichotomic observables, A_2 and A_2', acting on the second subsystem. The measurement outcomes for each observable are $+1$ and -1.

The correlation function $E(\cdot,\cdot)$ represents the expected value of the product of the measurement outcomes for the corresponding observables. For example, $E(A_1, A_2)$ gives the average value of the product of the results obtained when measuring A_1 and A_2 on their respective subsystems.

Let the Bell-CHSH polynomial be defined as:

$$S_2 = E\left(A_1A_2\right) - E\left(A_1A_2'\right) + E\left(A_1'A_2\right) + E\left(A_1'A_2'\right), \tag{8}$$

where S_2 characterizes the correlations between measurements on the two subsystems.

If the system satisfies a local hidden variable model, the following CHSH inequality holds (Werner and Wolf 2001; Alsina et al. 2016):

$$|S_2| \leq 2.$$

In contrast, for a quantum two-qubit system, the maximum value that $|S_2|$ can achieve by a quantum state is $2\sqrt{2}$, a result known as Tsirelson's bound (Cirel'son 1980; Tsirel'son 1987). Consequently, quantum mechanics violates local hidden variables theory.

For a three-partite system, the CHSH framework is generalized by the Mermin's polynomial defined as follows:

$$\begin{aligned} M_3 = {} & E\left(A_1A_2A_3'\right) + E\left(A_1A_2'A_3\right) + E\left(A_1'A_2A_3\right) \\ & - E\left(A_1'A_2'A_3'\right), \end{aligned} \tag{9}$$

where $E(\cdot,\cdot,\cdot)$ is the correlation function for three observables.

Under a local hidden variable model, Mermin's inequality holds:

$$|M_3| \leq 2.$$

For a quantum three-qubit system, the maximum value of $|M_3|$ is 4.

To identify genuine three-partite non-locality (Svetlichny (1987)), Svetlichny introduced a stronger set of inequalities. The Svetlichny polynomial is defined as:

$$\begin{aligned} S_3 = {} & E\left(A_1A_2A_3\right) + E\left(A_1A_2A_3'\right) + E\left(A_1A_2'A_3\right) + E\left(A_1'A_2A_3\right) \\ & - E\left(A_1'A_2'A_3'\right) - E\left(A_1'A_2'A_3\right) - E\left(A_1'A_2A_3'\right) - E\left(A_1A_2'A_3'\right). \end{aligned} \tag{10}$$

For a system satisfying a hybrid hidden variable model, the following Svetlichny inequality holds:

$$|S_3| \leq 4.$$

In the quantum case, the maximum violation achievable is $|S_3| = 4\sqrt{2}$.

It is important to note that a violation of Mermin's inequality does not necessarily imply genuine multipartite non-locality, as it can arise from strong bipartite correlations. In contrast, a violation of the Svetlichny inequality unambiguously reveals genuine three-partite non-locality.

5.1 Experimental results

The previous part of this work has been conducted on systems ranging from 2 to 5 qubits. However, let's note that the non-locality analysis has only been performed for 2 and 3 qubits. As shown in the recent study by Granda et al. (2024), increasing the number of qubits makes it much harder to obtain states that violate the Svetlichny inequality. This challenge can be alleviated by constructing a balanced dataset. However, achieving this requires an excessive number of states to generate a sufficiently large and representative set of samples for each class. A similar challenge arises with Mermin inequalities, though in the opposite sense: it is actually quite difficult to find states that **do not** violate

the inequality. Additionally, given that we are dealing with a binary classification problem, in this case, we have included a comparison with another quantum-inspired classifier, the Helstrom Quantum Classifier (HQC) (Helstrom 1969; Sergioli et al. 2019). Like in the previous section, we employed a random state generator for each class. Again, each experiment involves datasets of varying sizes (10k, 15k, 20k, 25k and 30k). In each instance, we used an 80% training set (with grid search 5-fold validation) and a 20% test set partition.

In this section, we present the numerical results obtained from our experiments evaluating the performance of the PGM Classifier with **two tensor copies** compared to various standard classifiers in detecting non-locality using the **CHSH inequality** for 2-qubit systems (5.1.1), **Mermin inequality** (5.1.2) and **Svetlichny inequality** (5.1.3) for 3-qubit systems. The setup of this experiment follows what has been described at the beginning of Subsection 4.1.

The average balanced accuracy for each classifier for all samples presented in Table 9, 10 and 11 demonstrate strong performances of the PGM Classifier across different datasets. Individual heatmaps illustrating the balanced accuracy achieved by the best-trained model of each classifier on each dataset are presented in Figs. 14, 17 and 20.

Additionally, the biclustering heatmaps shown in Figs. 15, 18 and 21, comparing classifiers for each dataset, indicate that the average winning ratio of the PGM Classifier exceeds that of most of all other classifiers. Finally, the confusion matrices shown in Figs. 16, 19 and 22 provide further insights into the classifiers' performance by illustrating the detailed distribution of true positive, false positive, true negative, and false negative predictions. They complement the balanced accuracy metric by highlighting the classifiers' ability to correctly identify non-local samples.

5.1.1 2-Qubit Case: CHSH inequality

Table 9 Average Balanced Accuracy for each classifier

Classifier	Average Balanced Accuracy
Support Vector Machine (SVM)	0.901 ± 0.016
MLP	0.875 ± 0.024
PGM	0.844 ± 0.011
Nearest Neighbors (NN)	0.839 ± 0.016
Random Forest (RF)	0.802 ± 0.032
Helstrom Bound Classifier	0.747 ± 0.007
Extra Tree (ET)	0.735 ± 0.028
Quadratic Discriminant Analysis (QDA)	0.503 ± 0.004
Nearest Centroid (NC)	0.503 ± 0.003
Dummy Classifier (Dummy)	0.501 ± 0.001
Linear Discriminant Analysis (LDA)	0.496 ± 0.008
Bernoulli Naive Bayes (B-NB)	0.495 ± 0.006
Gaussian Naive Bayes (G-NB)	0.495 ± 0.008

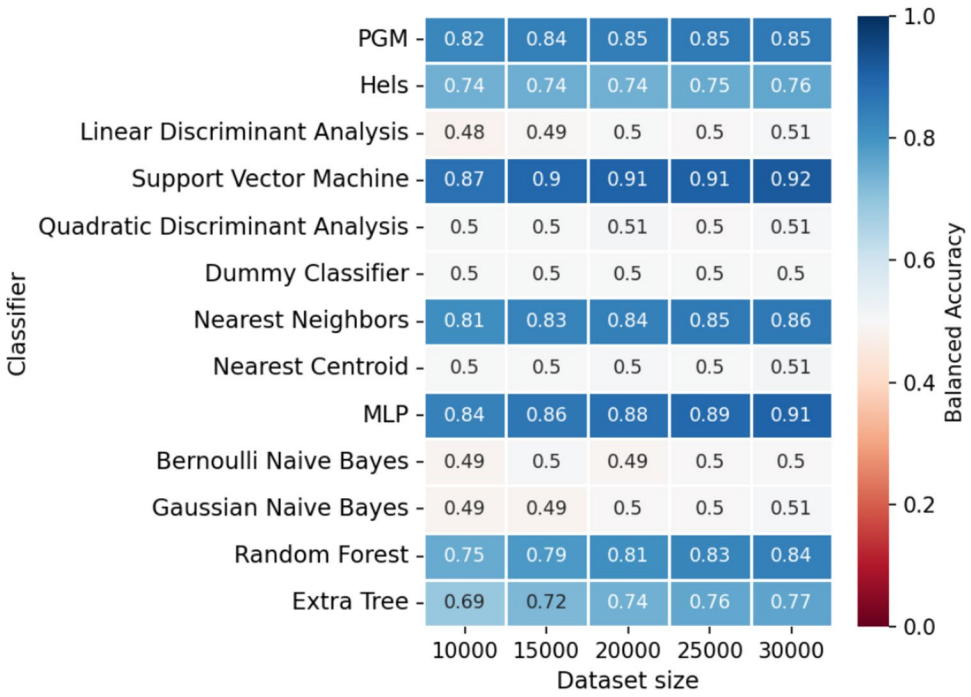

Fig. 14 Heatmap showing the Balanced Accuracy for different classifiers in the CHSH case, as a function of the number of samples

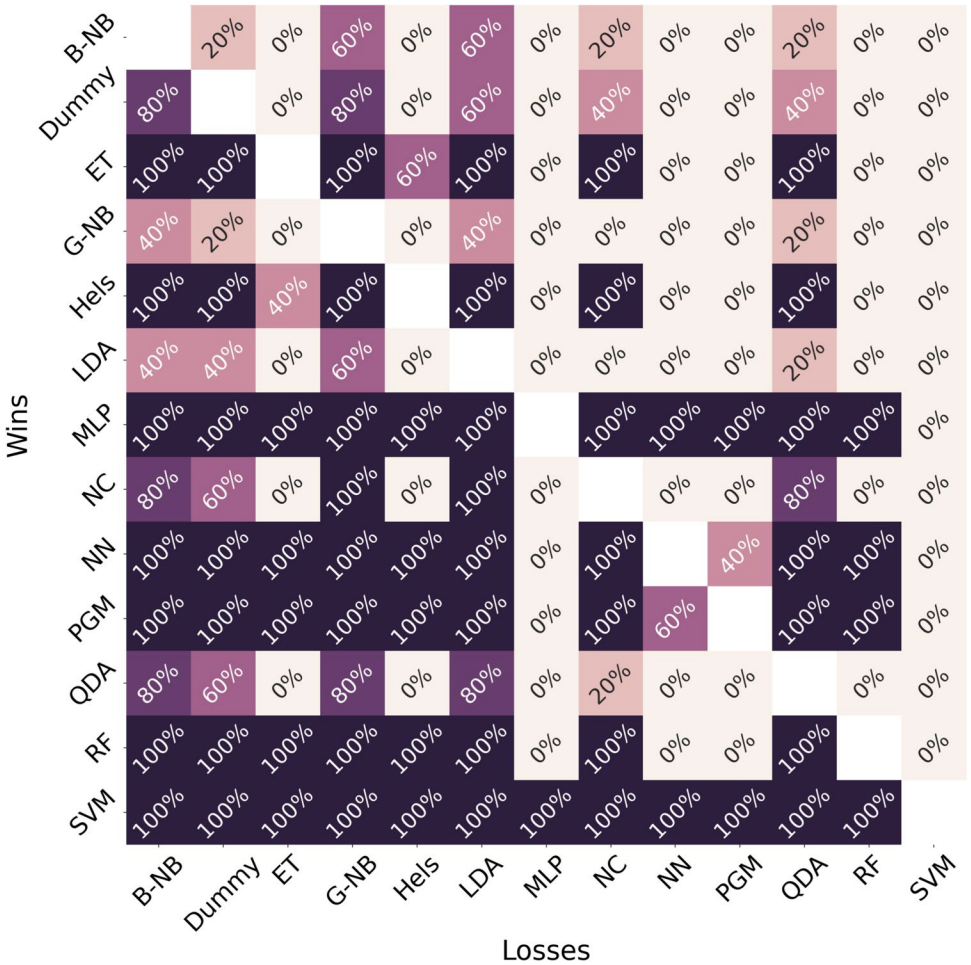

Fig. 15 Biclustering map, when a classifier A outperforms a classifier B according to the Balanced Accuracy. A lighter color indicates a lower percentage

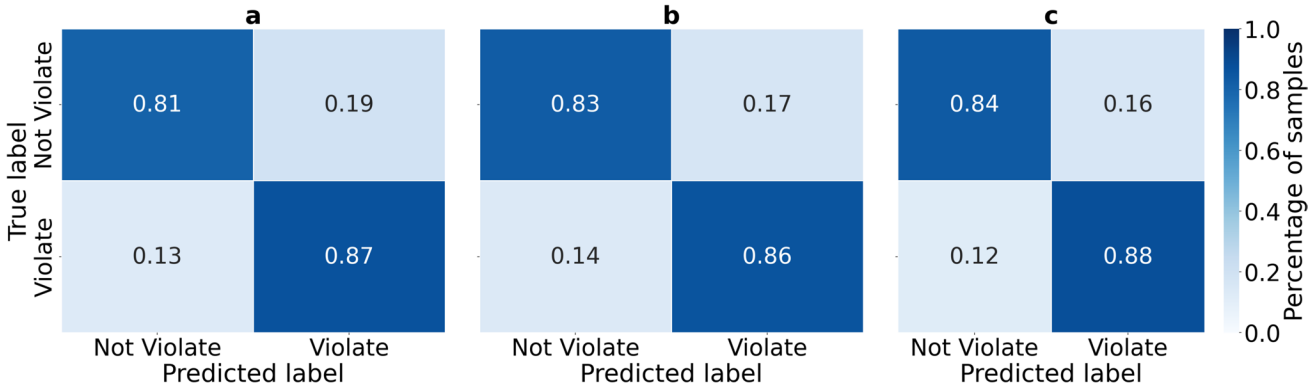

Fig. 16 Confusion matrix for a 2-qubit system for: (a) 10,000, (b) 20,000, and (c) 30,000 samples

5.1.2 3-Qubit Case: Mermin inequality

Table 10 Average Balanced Accuracy for each classifier

Classifier	Average Balanced Accuracy
Support Vector Machine (SVM)	0.795 ± 0.033
PGM	0.782 ± 0.016
Helstrom Bound Classifier	0.743 ± 0.013
Nearest Neighbors (NN)	0.739 ± 0.018
Random Forest (RF)	0.685 ± 0.022
MLP	0.646 ± 0.040
Extra Tree (ET)	0.643 ± 0.018
Quadratic Discriminant Analysis (QDA)	0.503 ± 0.003
Dummy Classifier (Dummy)	0.503 ± 0.001
Nearest Centroid (NC)	0.503 ± 0.003
Gaussian Naive Bayes (G-NB)	0.501 ± 0.004
Bernoulli Naive Bayes (B-NB)	0.500 ± 0.004
Linear Discriminant Analysis (LDA)	0.500 ± 0.003

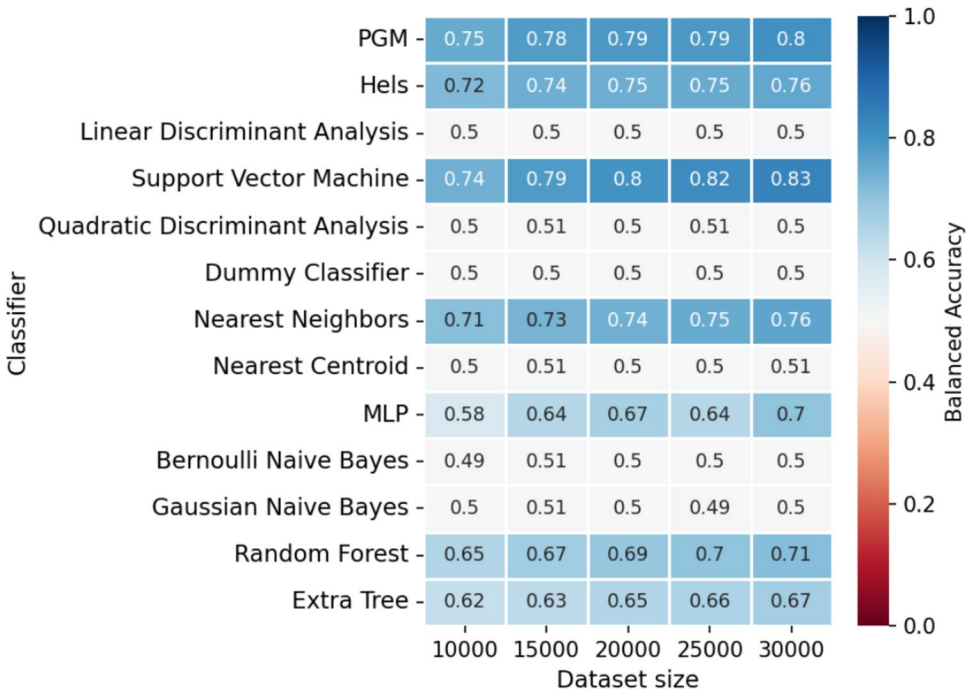

Fig. 17 Heatmap showing the Balanced Accuracy of different classifiers in the Mermin inequality case, as a function of the number of samples

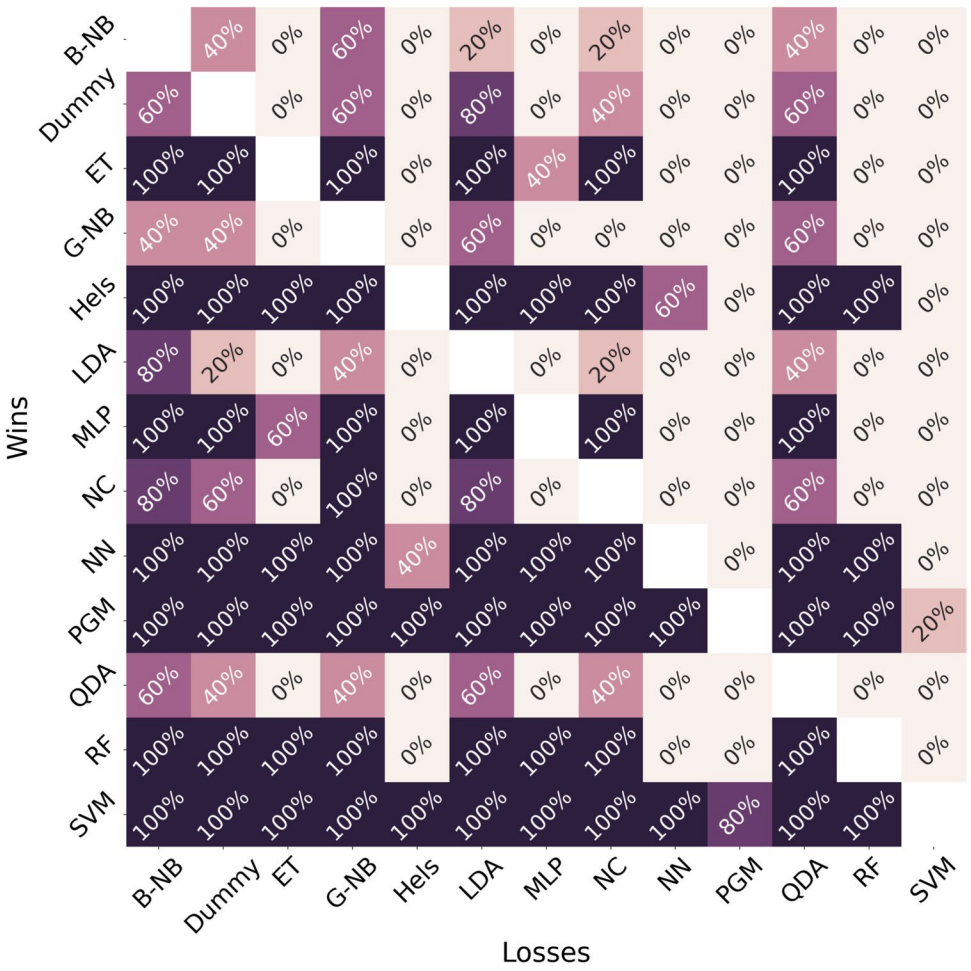

Fig. 18 Biclustering map, when a classifier A outperforms a classifier B according to the Balanced Accuracy. A lighter color indicates a lower percentage

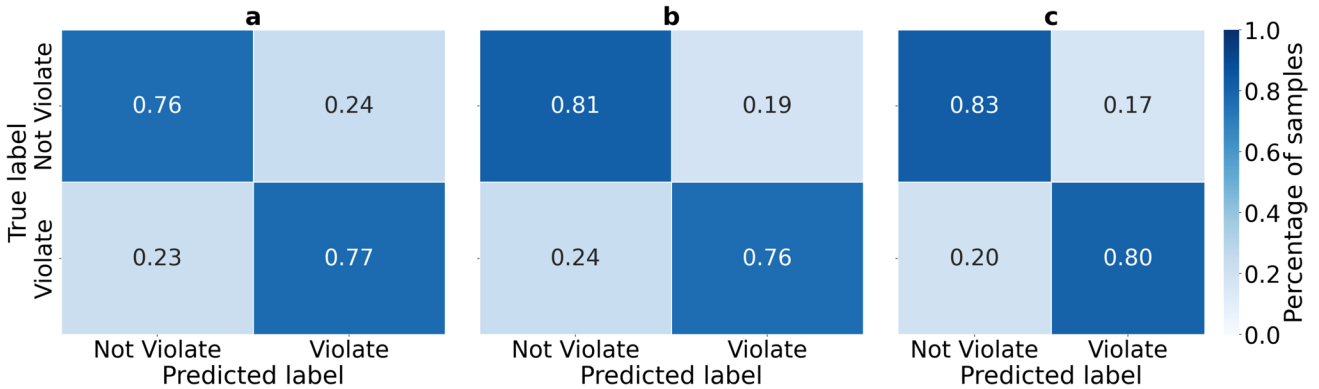

Fig. 19 Confusion matrix for a 3-qubit system for: (a) 10,000, (b) 20,000, and (c) 30,000 samples

5.1.3 3-Qubit Case: Svetlichny inequality

Table 11 Average Balanced Accuracy for each classifier

Classifier	Average Balanced Accuracy
Support Vector Machine (SVM)	0.909 ± 0.044
PGM	0.849 ± 0.034
Random Forest (RF)	0.842 ± 0.067
Helstrom Bound Classifier	0.836 ± 0.040
MLP	0.829 ± 0.072
Nearest Neighbors (NN)	0.818 ± 0.044
Extra Tree (ET)	0.776 ± 0.054
Quadratic Discriminant Analysis (QDA)	0.533 ± 0.054
Gaussian Naive Bayes (G-NB)	0.517 ± 0.025
Nearest Centroid (NC)	0.517 ± 0.025
Linear Discriminant Analysis (LDA)	0.516 ± 0.029
Bernoulli Naive Bayes (B-NB)	0.509 ± 0.018
Dummy Classifier (Dummy)	0.503 ± 0.001

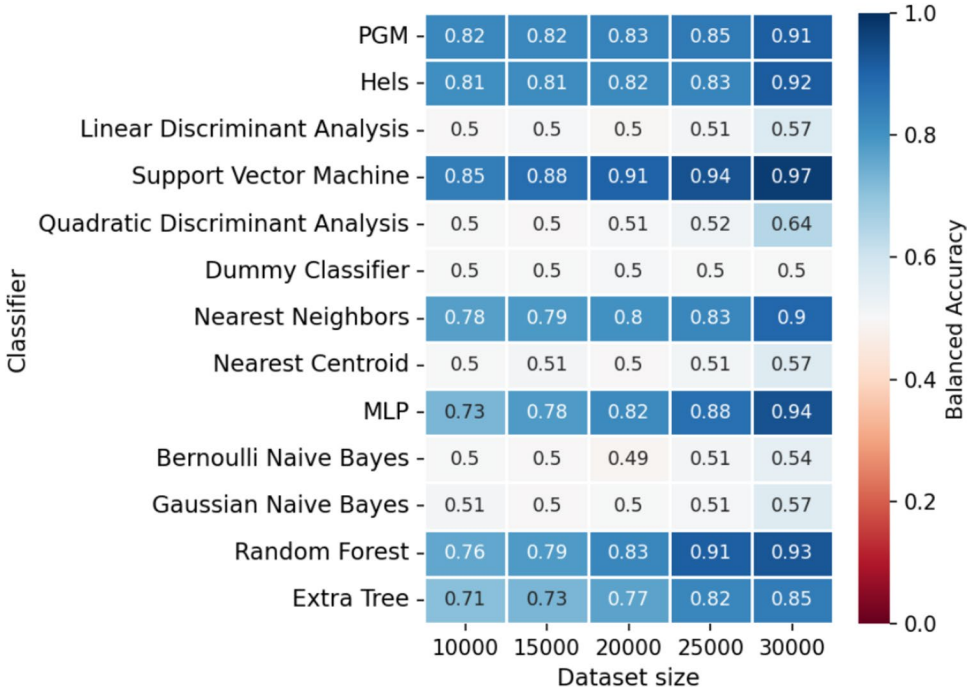

Fig. 20 Heatmap showing the Balanced Accuracy of different classifiers in the Svetlichny inequality case, as a function of the number of samples

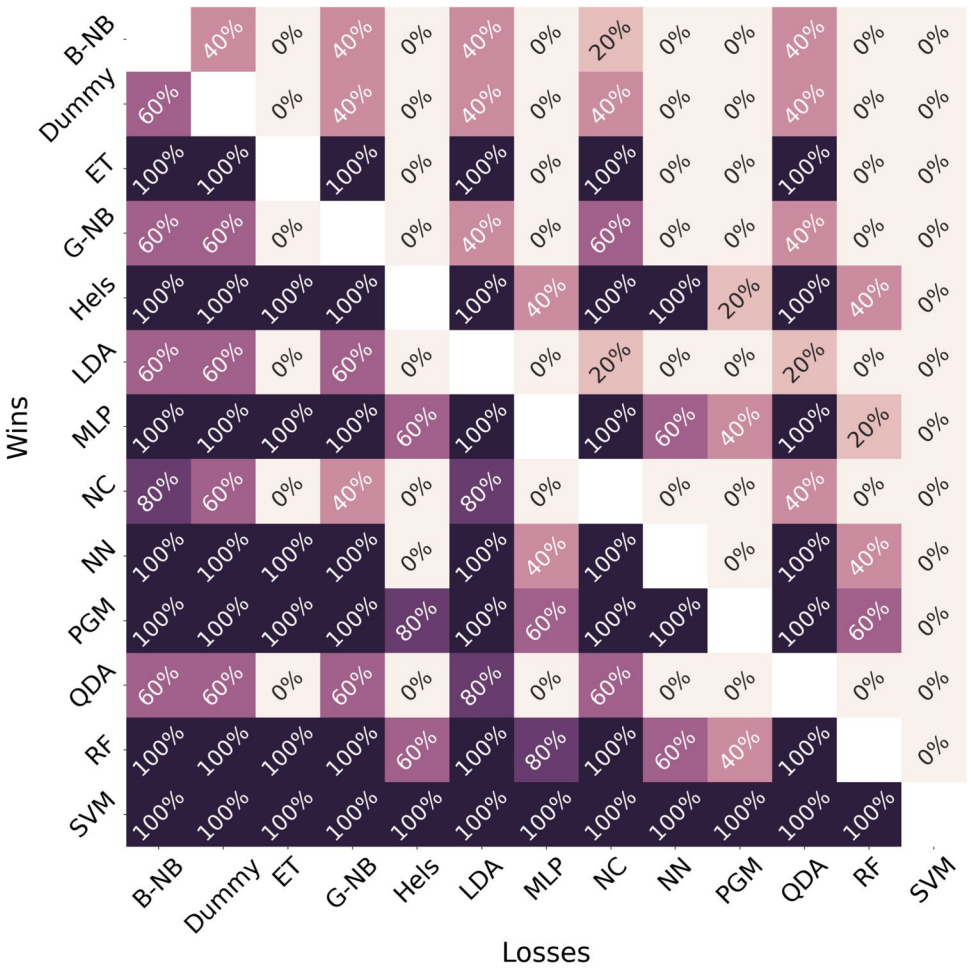

Fig. 21 Biclustering map, when a classifier A outperforms a classifier B according to the Balanced Accuracy. A lighter color indicates a lower percentage

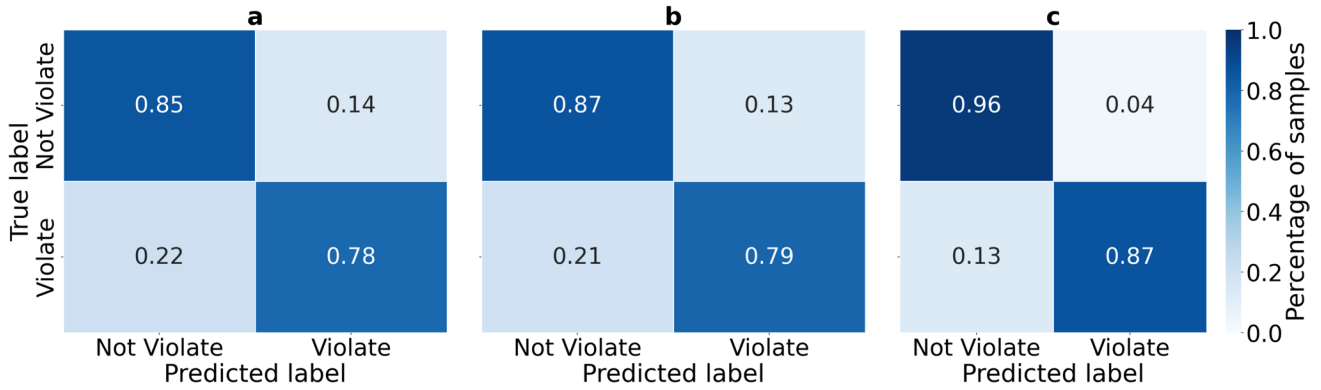

Fig. 22 Confusion matrix for a 3-qubit system for: (a) 10,000, (b) 20,000, and (c) 30,000 samples

6 Discussion

In the introduction of this article, we discussed how the classification of quantum states is a particularly compelling topic and how several approaches have already been developed to address this challenge. In this work, we propose an approach based on quantum state discrimination, and in particular, on the Pretty Good Measurement. The motivation for employing a PGM-based quantum classification framework can be systematically examined across three interrelated domains: foundational principles, computational advantages, and practical implementability (17, 18, 19, 20, 21, 22, and 23).

At the foundational and theoretical level, the motivation to employ a PGM-based classification stems primarily from the intrinsic compatibility between the formal structure of quantum state discrimination and the distinctive features of quantum correlations, such as entanglement. Unlike classical classification methods–which are typically grounded in geometric separability criteria in feature space or heuristic optimization strategies –the PGM-inspired quantum classifier is rooted in principles native to quantum state discrimination. As a result, it naturally and coherently incorporates the mathematical structures associated with entangled states, effectively capturing and leveraging features that are often overlooked by conventional classical methods.

From a computational standpoint, the PGM classifier offers a significant simplification over standard multiclass classification approaches. Classical techniques generally handle multiclass discrimination through combinatorial strategies based on binary schemes, such as One-vs-Rest or One-vs-One, which introduce additional complexity and interpretative inconsistencies. By contrast, the PGM-based classifier inherently manages multiclass discrimination in a unified, single-step operation, thereby reducing both computational and interpretative complexity, as shown in Giuntini et al. (2023b).

Finally, from a practical implementation perspective, recent advances point toward the feasibility of a fully quantum realization of the PGM classifier. In a previous study (Giuntini et al. 2023a), it was demonstrated that, by leveraging the Naimark extension theorem, any Positive Operator-Valued Measure (POVM)–including a PGM–can be operationally represented through sequences of unitary operators implementable on quantum circuits. That same work also introduced a prototype quantum circuit capable of performing the testing phase of the PGM-based classifier. Moreover, an important complementary direction concerns the training phase, which typically involves the computation of the Moore-Penrose pseudoinverse–often the primary computational bottleneck. Recent developments in quantum algorithms, particularly those based on the Singular Value Transformation (SVT) framework, offer a promising solution (Rebentrost et al. 2018; Gilyén et al. 2019). This method, building on and generalizing Hamiltonian simulation techniques, enables the efficient application of polynomial transformations to the singular values of a block-encoded matrix within a unitary operator. Notably, it allows for the pseudoinverse to be implemented with exponential precision, using quantum circuits that are

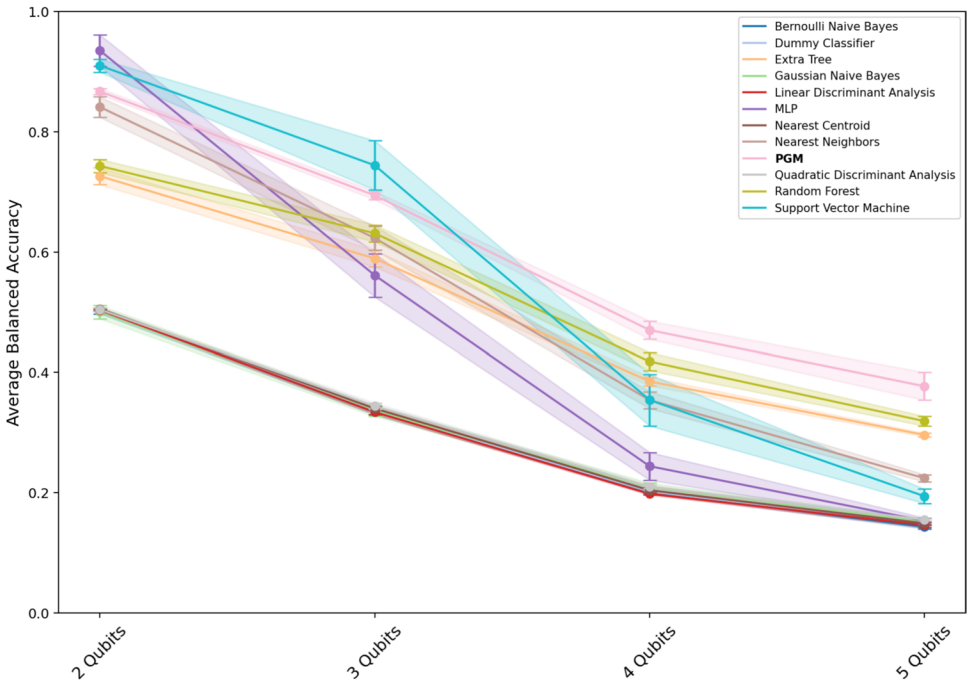

Fig. 23 Comparison of entanglement/separabilty classification performances for different numbers of qubits. The graph shows the average balanced accuracy of various classifiers. Error bars represent the standard deviation, and shaded regions indicate confidence intervals for each classifier

structurally simple, resource-efficient (e.g., requiring only a constant number of ancilla qubits), and often near-optimal in complexity. This framework has already been applied to quantum versions of principal component regression and other machine learning tasks, suggesting that the core linear algebraic operations required during PGM classifier training can likewise be executed within this unified and scalable quantum algorithmic paradigm. These developments will be considered for a future work and open the door to a fully quantum implementation of both the training and testing phases, potentially unlocking the full computational advantages offered by quantum machine learning.

The idea of this work is to show how a supervised quantum-inspired model, based on quantum state discrimination, can capture physical aspects related to quantum correlation by analyzing solely the configuration of the density operator under consideration in each instance.

The results presented in this study clearly highlight that classifying quantum correlations is, in general, a highly challenging task. While PGM classifier (together with most of the other classifiers) exhibit very good performance for a small number of qubits (2 or 3 qubits), their accuracy deteriorates as the number of qubits increases. At first glance, this can be naturally attributed to the fact that an increase in the number of qubits inevitably leads to a proportional rise in the number of classes. Specifically, for our PGM classifier, we observe that fine-grained classification among different types of separable states is particularly challenging.

Notably, the data reveal that the classification of factorized and entangled states remains acceptable for systems of 4 or 5 qubits, whereas the statistical significance of the classification among various classes of separable states diminishes significantly as early as the 4-qubit case. The difficulty in detecting quantum separability compared to factorization and entanglement is perhaps unsurprising. However, our experiments with the PGM classifier reveal additional insights that could inform future research. First, it was observed that classification accuracy benefits from an increased number of samples. Yet, this improvement is largely confined to the factorized and entangled state classes, while the classification of separable states shows limited enhancement even with more samples. Moreover, within the separable state category, the distribution of false positives is notably uneven. For instance, in the 5-qubit case (30,000 samples), type-1 and type-2 separability is frequently misclassified as entanglement, whereas type-4 and type-5 separability is more often confused with factorization. This trend suggests intriguing avenues for further investigation and will be thoroughly explored in future work focusing exclusively on the classification of separable states.

Finally, it is worth noting that, whether for few-qubit cases–where classification performs well–or for larger systems–where all classifiers struggle–the PGM consistently ranks among the top-performing methods compared to most of the classical classifiers (as shown in Fig. 23). Thus, while the classification of quantum correlations, especially for separable states, remains a particularly challenging task, we believe that our findings provide valuable insights for future studies and further establish the PGM as a highly promising quantum-inspired classifier.

Regarding the classification of non-locality, experiments performed for two-qubit systems (classification of CHSH inequality violations) show a high degree of accuracy, reaffirming the effectiveness of the PGM in distinguishing nonlocal correlations from local ones. Extending the analysis to three-qubit systems (Mermin and Svetlichny inequalities violation classification), the results indicate a consistency in which the PGM maintains a strong outperformance over most of the other classifiers.

Notably, the PGM achieves better performance for the Svetlichny inequality compared to the Mermin inequality. This observation can be explained by the fundamental difference between these two inequalities. A violation of Mermin's inequality does not necessarily imply genuine multipartite non-locality, as it can arise from strong bipartite correlations. In contrast, a violation of the Svetlichny inequality unambiguously reveals genuine three-partite nonlocality. Consequently, the Svetlichny inequality imposes a more stringent condition that contributes to the superior performance of the classifier in identifying non-locality.

However, the challenge of generating sufficient non-local states for systems of more qubits in the case of Svetlichny inequality remains significant (creation of random states), emphasizing the need for more efficient generation techniques or alternative approaches to handle higher dimensional non-local correlations.

Appendix: Tables with hyperparameter space and optimized hyperparameters for each classifier

At the beginning of Section 4.1, the experimental setup for all the experiments conducted throughout the paper was described in detail. This appendix reports the tables related to the hyperparameter search space for each classifier (12), as well as the tables showing the best configurations, both for the Factorized/Separable/Entangled classification task (13) and for the Local/Non-Local classification task (14).

Table 12 Hyperparameter space employed for grid search in the classification tasks of factorized, separable, and entangled quantum states, as well as in the identification of non-local quantum correlations

Classifier	Hyperparameter Space
LDA	solver = {svd, lsqr, eigen}
QDA	reg_param = {0.1, 0.5, 1.0}, store_covariance = {True, False}, tol = {1e-4, 1e-3, 1e-2}
Dummy	strategy = {stratified, most_frequent, prior, uniform, constant}, constant = {0, 1}
SVM	C = {0.1, 1, 10}, kernel = {linear, poly, rbf}, degree = 3, gamma = {scale, 0.001, 0.01, 0.1, 1}
Nearest Neighbor	n_neighbors = {3, 5, 7, 10}, weights = {uniform, distance}, leaf_size = {10, 30, 50}, p = {1, 2}
Nearest Centroid	metric = {euclidean, manhattan}, shrink_threshold = {None, 0.1, 0.5, 1.0}
MLP	activation = relu, solver = adam, learning_rate_init = {0.1, 0.01, 0.001, 0.0001}, max_iter = 3000, early_stopping = False, hidden_layer_sizes = various configurations
Bernoulli NB	alpha = 1
Gaussian NB	priors = None
Random Forest	min_samples_split = {3, 5, 10}, n_estimators = {100, 300}, max_depth = {3, 5, 15, 25}
Extra Trees	n_estimators = {50, 75, 100, 125}, min_samples_leaf = {20–45}, min_samples_split = {15–35}
PGM	rescale = 0.5, 1
HELS	rescale = 0.5, 1

Table 13 Optimized parameters for the classification of factorized, separable, and entangled quantum states (2–5 Qubits, 30,000-State Dataset)

Classifier	2 Qubits	3 Qubits	4 Qubits	5 Qubits
PGM	rescale: 0.5			
Hels	rescale: 0.5	X		
Linear Discriminant Analysis	solver: `svd`			
Support Vector Machine	C: 10, degree: 3, gamma: `scale`, kernel: `rbf`	C: 10, degree: 3, gamma: 1, kernel: `rbf`	C: 10, degree: 3, gamma: `scale`, kernel: `rbf`	
Quadratic Discriminant Analysis	reg_param: 0.1, store_covariance: True, tol: 0.0001	reg_param: 0.0, store_covariance: True, tol: 0.0001		
Dummy Classifier	strategy: `uniform`	strategy: `uniform`	strategy: `most_frequent`	strategy: `stratified`
Nearest Neighbors	leaf_size: 10, n_neighbors: 10, p: 2, weights: `distance`	leaf_size: 10, n_neighbors: 5, p: 2, weights: `distance`	leaf_size: 10, n_neighbors: 10, p: 1, weights: `distance`	leaf_size: 10, n_neighbors: 3, p: 1, weights: `distance`
Nearest Centroid	metric: `manhattan`, shrink: 0.1	metric: `euclidean`, shrink: 0.5	metric: `manhattan`, shrink: 1.0	metric: `manhattan`, shrink: None
MLP		activation: `relu`, early_stopping: `False`, hidden_layer_sizes: [32, 64×13, 32],lr=0.001, solver: `adam`	activation: `relu`, early_stopping: `False`, hidden_layer_sizes: [8, 16, 16, 8], lr=0.01, solver: `adam`	activation: `relu`, early_stopping: `False`, hidden_layer_sizes: [32, 64×13, 32], lr=0.001, solver: `adam`
Bernoulli Naive Bayes	alpha: 1			
Gaussian Naive Bayes	priors: None			
Random Forest	split: 3			split: 5
Extra Tree	estimators: 125			estimators: 100

Table 14 Optimized parameters for the classification of non-local quantum states using different inequalities (2–3 qubits, 30,000-state dataset)

Classifier	2Q CHSH	3Q Svetlichny	3Q Mermin
PGM	rescale: 0.5		
Hels	rescale: 0.5	X	X
Linear Discriminant Analysis	solver: `svd`		
Support Vector Machine	C: 10, degree: 3, gamma: `scale`, kernel: `rbf`		C: 10, degree: 3, gamma: 1, kernel: `rbf`
Quadratic Discriminant Analysis	reg_param: 1.0, store_covariance: True, tol: 0.0001	reg_param: 0.0, store_covariance: True, tol: 0.0001	reg_param: 0.0, store_covariance: True, tol: 0.0001
Dummy Classifier	constant: 0, strategy: most_frequent	strategy: `uniform`	strategy: `uniform`
Nearest Neighbors	leaf_size: 10, n_neighbors: 10, p: 2, weights: `distance`	p: 1, weights: `uniform`	p: 2, weights: `distance`
Nearest Centroid	metric: `euclidean`, shrink_threshold: None		
MLP		activation: `relu`, early_stopping: `False`, hidden_layer_sizes: [32, 64×14, 32], learning_rate_init: 0.001, max_iter: 3000, solver: `adam`	activation: `relu`, early_stopping: `False`, hidden_layer_sizes: [8, 16, 16, 8], learning_rate_init: 0.01, solver: `adam`
Bernoulli Naive Bayes	alpha: 1		
Gaussian Naive Bayes	priors: None		
Random Forest	max_depth: 25, min_samples_split: 3, n_estimators: 300	n_estimators: 100	max_depth: 25, min_samples_split: 3, n_estimators: 300
Extra Tree	min_samples_leaf: 20, min_samples_split: 15, n_estimators: 125	n_estimators: 50	min_samples_leaf: 20, min_samples_split: 15, n_estimators: 125

Acknowledgements Roberto Giuntini and Giuseppe Sergioli are partially supported by the projects:
- "CORTEX The Cost of Reasoning: Theory and Experiments", funded by the Ministry of University and Research (Prin 2022, cod. 2022ZLLR3T).
- "Quantum Models for Logic, Computation and Natural Processes (Qm4Np)" funded by the Ministry of University and Research (Prin-Pnrr 2022 cod. P2022A52CR).
- "An algebraic approach to hyperintensionality", funded by Fondazione di Sardegna (CUP F23C25000360007).

Roberto Giuntini is partially funded by the TÜV SÜD Foundation, the Federal Ministry of Education and Research (BMBF) and the Free State of Bavaria under the Excellence Strategy of the Federal Government and the Länder, as well as by the Technical University of Munich-Institute for Advanced Study.
Carla Rieger is supported by CERN through the CERN Quantum Technology Initiative and by the Wolfgang Gentner Programme of the German Federal Ministry of Education and Research (grant no. 13E18CHA).

Author Contributions All authors contributed equally to the conceptualization, experimental work, and manuscript preparation. Each author played an integral role in the development of the research questions, the design and implementation of the experiments, and the analysis and interpretation of the results. The drafting and revision of the manuscript were conducted collaboratively, ensuring that all sections reflect the collective insights and expertise of the authors. All authors have reviewed and approved the final version of the manuscript.

Funding Open access funding provided by Università degli Studi di Cagliari within the CRUI-CARE Agreement.

Code availability The code for generating and sampling pure quantum states is publicly accessible at https://github.com/QuantumUnica/Pure_QState_Generator. Additionally, the code and data partitions required to replicate the experiments are provided at https://github.com/QuantumUnica/Classification-of-Pure-Quantum-Correlations.

Data Availability All the experimental data reported in the manuscript are reproducible by running the codes available at the following links: https://github.com/QuantumUnica/Pure_QState_Generator, https://github.com/QuantumUnica/Classification-of-Pure-Quantum-Correlations

Declarations

Competing interests The authors declare no competing interests.

References

Alsina D, Cervera A, Goyeneche D, Latorre JI, Życzkowski K (Sep2016) Operational approach to bell inequalities: Application to qutrits. Phys Rev A 94(3):032102. https://doi.org/10.1103/PhysRevA.94.032102

Bell JS (1964) On the einstein podolsky rosen paradox. Physics Physique Fizika 1:195–200. https://doi.org/10.1103/PhysicsPhysiqueFizika.1.195

Brunner N, Cavalcanti D, Pironio S, Scarani V, Wehner S (2014) Bell nonlocality. Rev Mod Phys 86:419–478. https://doi.org/10.1103/RevModPhys.86.419

Cirel'son BS (1980) Quantum generalizations of bell's inequality. Lett Math Phys 4(2):93–100. ISSN 1573–0530. https://doi.org/10.1007/BF00417500

Clauser JF, Horne MA, Shimony A, Holt RA (Oct1969) Proposed experiment to test local hidden-variable theories. Phys Rev Lett 23:880–884. https://doi.org/10.1103/PhysRevLett.23.880

El Ayachi F, El Baz M (2022) General classification of entanglement using machine learning. arXiv:2210.07711

Gilyén A, Su Y, Low GH, Wiebe N (2019) Quantum singular value transformation and beyond: exponential improvements for quantum matrix arithmetics. In Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing, STOC 2019, page 193–204, New York, NY, USA. Association for Computing Machinery. ISBN 9781450367059. https://doi.org/10.1145/3313276.3316366

Giuntini R, Granda AC, Freytes H, Holik FH, Sergioli G (2023) Multi-class classification based on quantum state discrimination. Fuzzy Sets Syst 467:108509. https://doi.org/10.1016/j.fss.2023.03.012

Giuntini R, Holik FH, Park DK, Freytes H, Blank C, Sergioli G (2023b) Quantum-inspired algorithm for direct multi-class classification. Applied Soft Computing 134: 109956. ISSN 1568–4946. https://doi.org/10.1016/j.asoc.2022.109956

Granda AC, Holik FH, Sergioli G, Giuntini R, Freytes H (2024) Geometrical aspects of resources distribution in quantum random circuits. arXiv:2405.01650

Harney C, Pirandola S, Ferraro A, Paternostro M (Apr2020) Entanglement classification via neural network quantum states. New J Phys 22(4):045001. https://doi.org/10.1088/1367-2630/ab783d

Hausladen P, Wootters WK (1994) A 'pretty good' measurement for distinguishing quantum states. J Mod Opt 41(12):2385–2390. https://doi.org/10.1080/09500349414552221

Helstrom CW (1969) Quantum detection and estimation theory. J Stat Phys 1 (2): 231–252. ISSN 1572–9613. https://doi.org/10.1007/BF01007479

Lin X, Chen Z, Wei Z (2023) Quantifying quantum entanglement via a hybrid quantum-classical machine learning framework. Phys Rev A 107:062409. https://doi.org/10.1103/PhysRevA.107.062409

Barnett SM, Croke S (2009) Quantum state discrimination Adv Opt Photon 1(2):238–278. https://doi.org/10.1364/AOP.1.000238

Ma YC, Yung MH (2018) Transforming bell's inequalities into state classifiers with machine learning. npj Quantum Information 4(1):34. ISSN 2056-6387. https://doi.org/10.1038/s41534-018-0081-3

Mengoni R, Di Piero A (2019) Kernel methods in quantum machine learning. Quantum Machine Intelligence 1 (3): 65–71. ISSN 2524-4914. https://doi.org/10.1007/s42484-019-00007-4

Mermin ND (1990) Extreme quantum entanglement in a superposition of macroscopically distinct states. Phys Rev Lett 65(15):1838–1840. https://doi.org/10.1103/PhysRevLett.65.1838

Mochon C (2006) Family of generalized "pretty good" measurements and the minimal-error pure-state discrimination problems for which they are optimal. Phys Rev A 73:032328. https://doi.org/10.1103/PhysRevA.73.032328

Pawłowski J, Krawczyk M (2024) Identification of quantum entanglement with siamese convolutional neural networks and semi-supervised learning. arXiv:2210.07410

Rebentrost P, Steffens A, Marvian I, Lloyd S (2018) Quantum singular-value decomposition of nonsparse low-rank matrices. Phys Rev A 97:012327. https://doi.org/10.1103/PhysRevA.97.012327

Santucci E, Sergioli G (2018) Classification Problem in a Quantum Framework, pages 215–228. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-319-74971-6_16

Schatzki L, Arrasmith A, Coles PJ, Cerezo M (2021) Entangled datasets for quantum machine learning. https://arxiv.org/abs/2109.03400

Seevinck M, Svetlichny G (2002) Bell-type inequalities for partial separability in n-particle systems and quantum mechanical violations. Phys Rev Lett 89:060401. https://doi.org/10.1103/PhysRevLett.89.060401

Sergioli G (2020) Quantum and quantum-like machine learning: a note on differences and similarities. Soft Computing, 24 (14): 10247–10255. ISSN 1433-7479. https://doi.org/10.1007/s00500-019-04429-x

Sergioli G, Bosyk G, Santucci E, Giuntini R (2017) A quantum-inspired version of the classification problem. Int J Theor Phys 56:12. https://doi.org/10.1007/s10773-017-3371-1

Sergioli G, Giuntini R, Freytes H (2019) A new quantum approach to binary classification. PLoS ONE 14(5):1–14. https://doi.org/10.1371/journal.pone.0216224

Sirui L, Shilin H, Keren L, Jun L, Chen Jianxin L, Dawei JZ, Yi S, Duanlu Z, Bei Z (Jul2018) Separability-entanglement classifier via machine learning. Phys Rev A 98:012315. https://doi.org/10.1103/PhysRevA.98.012315

Svetlichny G (1987) Distinguishing three-body from two-body nonseparability by a bell-type inequality. Phys Rev D 35:3066–3069. https://doi.org/10.1103/PhysRevD.35.3066

Tsirel'son BS (1987) Quantum analogues of the bell inequalities. the case of two spatially separated domains. J Math Sci 36: 557

Vintskevich SV, Bao N, Nomerotski A, Stankus P, Grigoriev DA (2023) Classification of four-qubit entangled states via machine learning. Phys Rev A 107:032421. https://doi.org/10.1103/PhysRevA.107.032421

Werner RF, Wolf MM (2001) All-multipartite bell-correlation inequalities for two dichotomic observables per site. Phys Rev A 64(3):032112. https://doi.org/10.1103/PhysRevA.64.032112