The Journal of Supercomputing (2026) 82:234
https://doi.org/10.1007/s11227-026-08396-7

A framework to study the relationship between classical data and quantum characteristics in quantum machine learning

Iraitz Montalbán[1] · Elías F. Combarro[1] · José Ranilla[1]

Received: 16 October 2025 / Accepted: 20 February 2026

Abstract
Quantum Machine Learning (QML) holds the promise of improving conventional machine learning, but the conditions under which an advantage can be obtained are still unclear. Although many studies have been conducted in the literature, there is no conclusive evidence that allows us to determine which types of classical datasets or which problem complexities benefit from the use of quantum techniques. Moreover, there have been contradicting findings when dealing with small or unbalanced datasets. In order to systematically approach this challenge and clarify the need for quantum properties in specific tasks, we propose the adoption of a framework capable of exploring the solution space of quantum feature maps and providing solid evidence of their potential advantages. Our framework comprises three main components: (a) the construction of datasets designed to highlight the specific characteristics that are expected to be relevant in quantum machine learning tasks; (b) the definition of metrics that serve as proxies for quantum kernels with desirable properties; and (c) an evolutionary-guided search for quantum feature maps maximizing those metrics. When using this framework with sensible parameter choices, we obtain results suggesting that some previous studies may have reported overfitted outcomes. This shows that justifying the need for quantum mechanical properties might be beyond the actual scope of conventional classical tasks, since there is no clear quantum feature that contributes to the gain shown by some of the QML techniques most commonly applied in the literature.

Keywords Quantum machine learning · Kernel alignment · Quantum features

✉ José Ranilla
ranilla@uniovi.es

Iraitz Montalbán
UO311277@uniovi.es

Elías F. Combarro
efernandezca@uniovi.es

[1] Department of Computer Science, University of Oviedo, Gijón, Spain

Published online: 10 March 2026

1 Introduction

Quantum Machine Learning (QML) has emerged as a promising frontier at the intersection of quantum computing and machine learning [1]. By exploiting the principles of quantum mechanics—such as superposition and entanglement—QML aims to extend the representational power and computational efficiency of classical learning algorithms. Early studies have demonstrated potential advantages in specific contexts [2], yet the fundamental conditions under which QML truly outperforms classical methods remain an open question [3]. Despite significant theoretical and experimental progress, conclusive evidence on whether quantum models provide intrinsic learning benefits—beyond those achievable by classical counterparts—remains elusive.

One of the most discussed aspects of this debate lies in the role of quantum optimization and the extent to which it can enhance existing machine learning workflows. Quantum algorithms have shown competitive performance in optimization tasks [4], which are central to model training, feature selection [5], and hyperparameter tuning. However, these approaches often face scrutiny regarding scalability, stability, and robustness against noise. The hybridization of quantum and classical paradigms has thus become a pragmatic path forward, particularly in regimes where classical computational bottlenecks dominate [6].

A more fundamental question, however, concerns the representational advantage of quantum systems. Quantum feature maps and kernels define how classical data are embedded into high-dimensional Hilbert spaces, potentially enabling linear separability for data that are inseparable in the original space [2]. The inclusion of entanglement as a computational resource is often cited as a driver of this advantage, yet empirical evidence has been inconsistent [7, 8]. While some results indicate measurable improvements in classification accuracy when entangled embeddings are used, others find negligible or even detrimental effects. This variability highlights the need for systematic frameworks capable of isolating and quantifying the contribution of quantum characteristics—such as expressivity and entanglement—to learning performance.

Addressing this need, we propose a comprehensive framework designed to analyze the relationship between classical data complexity and quantum feature map characteristics. By engineering controlled datasets of tunable complexity and employing an evolutionary search over parametrized quantum circuits, our approach enables a structured investigation of how quantum embeddings adapt to and exploit classical data structures. The framework combines proxy metrics such as the Centered Kernel Alignment (CKA) with expressivity and entanglement measures to assess how quantum mechanical properties influence generalization and separability in classification tasks.

The remainder of this paper is structured as follows: Sect. 2 presents the theoretical background, introducing kernel-based learning and the role of quantum feature maps in data embedding. Section 3 details the proposed framework, including dataset construction, the Pauli-Genetic Embedding (PGE) mechanism that is central to our evolutionary search, and the theoretical bounds linking circuit depth to data complexity. Section 4 reports experimental results comparing classical and quantum kernels across datasets of varying complexity, highlighting the relationship between overfitting and quantum circuit expressivity. Finally, Sect. 5 summarizes the main findings

and outlines directions for future research, particularly regarding the conditions under which QML can yield verifiable and reproducible advantages over classical methods.

2 Related work

The fundamental capability of machine learning algorithms lies in their capacity to distinguish between distinct samples and establish mappings to target outputs—whether discrete class labels in classification or continuous values in regression. While the taxonomy of machine learning encompasses diverse methodological paradigms, the efficacy of most approaches fundamentally reduces to the separability of the data manifold, characterized by the existence and complexity of decision boundaries partitioning the feature space.

Consider the spectrum of separability complexity. At one extreme, linearly separable data admit a hyperplane $\mathrm{H} = \{x \in \mathbb{R}^n : \mathbf{w}^\top x + b = 0\}$ that perfectly partitions samples belonging to different classes—a configuration readily addressed by classical linear models. At the opposite extreme lies the pathological case where samples from distinct classes occupy identical positions in feature space, i.e., $\exists x_i, x_j$ such that $x_i = x_j$ but do not belong to the same class ($y_i \neq y_j$). This represents an inherently ill-posed problem admitting no deterministic solution, as any classifier must necessarily resort to probabilistic assignment. Between these boundary cases, there exists a continuum of intermediate complexities, the characterization of which is essential for assessing problem tractability.

The efficacy of kernel methods stems precisely from their ability to address this challenge through implicit feature space transformations. By employing a mapping $\phi : X \to \mathcal{H}$ to a (potentially infinite-dimensional) reproducing kernel Hilbert space, these methods project data into domains where linear separability may be achieved—effectively reducing complex nonlinear problems to linear ones in the transformed space.

To establish a rigorous framework for analyzing quantum machine learning techniques, we must therefore develop a bidirectional understanding: (1) quantitative assessment of intrinsic data complexity, including the geometry of class boundaries and the distribution of samples relative to decision surfaces, and (2) precise characterization of the representational capacity of feature maps—both classical and quantum—to induce favorable separability in their respective embedding spaces. This dual perspective provides the necessary foundation for systematic investigation of quantum embeddings prior to engaging with the additional technical complexities inherent to quantum computational models.

2.1 Mapping data into states

Kernel methods [9] have been widely used in traditional machine learning. This technique fundamentally embeds the original data into more expressive feature spaces, where complex dependencies with target variables become linearly separable. In this

representation, a suitable distance metric is required to define decision boundaries for classification.

The ability to represent data in a different dimensional space motivated the research line of quantum kernels for machine learning tasks [2, 10] tackling one of the first challenges when it comes to quantum models: classical data embedding. This approach was further explored by eminent researchers who empirically tested some of the potential advantages of these techniques. Works such as [8] even highlight, as a conclusion of the study, how kernels without entangling gates may behave better than those that include them. This is somewhat surprising, for it seems to remove part of the "quantumness" of the embedding process.

How classical data will be represented in the quantum space to ensure that the original learning objective is preserved is a challenging task on its own [11]. As mentioned in the discussion of [7], finding the right classical candidate problem and quantum embedding that demonstrate a benefit attributable solely to the quantum representation itself remains an open challenge. The latest advances approach the task of finding good quantum embedding techniques using heuristic methods and evolutionary algorithms that perform better than brute-force approaches but cannot guarantee the global optimality of the found embedding [12, 13].

To leverage quantum machine learning models, whether kernel-based or otherwise, the main task still resides in faithfully representing the original data from a classical feature space into a Hilbert feature space. The goal of this transformation between the two spaces is to maximize the separability between samples corresponding to various classes. The non-trivial task of finding the best quantum feature map is a complex optimization problem. It requires to find an encoding mechanism (or embedding) such that

$$|\psi(x)\rangle = U(x)|0\rangle^{\otimes n}$$

where U is the unitary transformation applied to x, a data sample. Variational parameters can be added to the U transformation so that $U(x, \theta)$ is also a common representation for these approaches.

Most approaches follow a pragmatic path, where quantum circuit composition is either determined using genetic algorithms [14, 15] or by selecting a fixed parametrized quantum circuit, which then requires fine-tuning to optimize free parameters and increase the separability between the samples [11]. This second option might be a better approach when a particular *a priori* knowledge exists on how to map the classical information using a specific quantum feature map configuration [8, 12]. In the absence of such knowledge and focusing on a theoretical machine that could execute any circuit without loss, an expressive representation is required to conduct a vast global search of all potential embeddings.

Once the embedding is obtained, a kernel can be created by evaluating the fidelity between two potential mapping states, such as in

$$K(x_i, x_j) = \mathrm{tr}[\rho(x_i)\rho(x_j)],$$

where $\rho(x) = |\psi(x)\rangle\langle\psi(x)|$. This motivates the use of the “fidelity kernel” name for this function. Even though projected quantum kernels [7] have been proposed as a

better-performing alternative, we have prioritized the fidelity kernel in this study as a direct path to understand the characteristics of the embeddings obtained, regardless of the final kernel composition technique.

Given that no *a priori* knowledge exists for the encoding in the general case, during this search we can provide a proxy so that the final classification task is not required to be solved each time in order to determine a suitable embedding. Previous works have leveraged this intermediate mechanism [13] that guides the search for the configuration that better splits samples belonging to different classes inside the Hilbert space. A common metric used in many of these previous studies is the Kernel Target Alignment (KTA) proposed in [11]. KTA is defined by

$$KTA(K, Y) = \frac{\langle K, Y \rangle_F}{\sqrt{\langle K, K \rangle_F \langle Y, Y \rangle_F}}, \tag{1}$$

where $\langle \cdot, \cdot \rangle_F$ represents the Frobenius inner product, K represents the kernel function over the X dataset, and $Y = yy^T$, where the components of y are the labels $y_n \in \{-1, 1\}$ representing the binary class for each X sample. Since the original metric did not correlate well with the actual generalization error of the model, a centered version of this metric was proposed in [16]. This second option was shown to be closer to the actual final performance obtained. Its expression is given by

$$CKA(K, Y) = KTA(HKH, HYH), \tag{2}$$

where $H = \mathbb{I} - (1/n)\mathbb{1} \cdot \mathbb{1}^T$ performs the mean subtraction of the matrix. Thus, the centering action focuses on the structural relationships between data points instead of magnitude and scaling effects, making this metric more robust for the iterative process described above.

2.2 Expressivity and entanglement

A key aspect that justifies this transition from classical to quantum data representation is that the complexity of the relationships in the classical feature space is potentially better expressed in the quantum domain. Therefore, for quantum kernels to offer an improvement, they must capture something that classical kernels do not. Thus, this ability to capture highly complex relationships present in the data should be linked to the expressivity of the embedding circuit and to how it can exploit entanglement as a resource. Quantifying entanglement for particular states using entanglement witnesses [17] or geometric metrics [18] may not be tractable when facing real hardware, as it requires full knowledge of the states being encoded. As an alternative, previous works have dealt with the measurement of this expressivity by relying on a finite set of samples [19, 20].

The ensemble of Haar random states has been previously used to evaluate the upper bound of expressivity for a given Parametrized Quantum Circuit (PQC). Such a quantum circuit can yield different binary strings upon measurement, depending on the values assigned to its parameters. Expressing the dissonance between the potential Haar-state measurement distribution and the outcome of a sufficiently large set of

randomly distributed parameters renders a metric that allows quantification of how fixed or flexible a particular quantum embedding might be. This relationship expressed in [20] is based on the KL divergence expressed

$$\text{Expr} = \mathrm{D}_{\mathrm{KL}}\left(\hat{\mathrm{P}}_{\mathrm{PQC}}(\mathrm{F};\theta)\,\|\mathrm{P}_{Haar}(\mathrm{F})\right), \tag{3}$$

where $\hat{\mathrm{P}}_{\mathrm{PQC}}$ is the estimated probability distribution of the fidelities resulting from the sampling statistics of the embedding circuit. Thus, a divergence of zero would mean the embedding can render any of the potential fidelity measures in the associated Hilbert space. It is important to notice that the lower this value, the higher the expressivity of the quantum embedding is as it poses the ability to represent any state in the target space.

For the entanglement capacity, similarly, the discrete sample-based version from [20] is defined by

$$\text{Ent} = \frac{1}{\|S\|}\sum_{\theta_i \in S} Q(|\psi_{\theta_i}\rangle) \tag{4}$$

where $S = \{\theta_i\}$ is the set of sampled circuit parameter vectors and Q stands for the Meyer and Wallach measure [21]. This gives a value of 0 when all measured states are product states and 1 when maximally entangled states are obtained. Alternative metrics like the Schmidt number can lead to more precise estimation, but they require a fully formed density matrix which for a parametrized circuit only accounts for one of the data points to be evaluated. For the sake of simplicity, in this work we have chosen the Meyer and Wallach approach to represent the entanglement capacity, using a subset of potential data points to estimate for the entanglement.

3 Proposed framework

In order to systematically study the effect of quantum properties when used for classical data classification tasks, we must be able to tweak and modify one potentially relevant aspect at a time and assess its impact. This requires a well-defined set of input datasets covering a broad range of cases, along with a systematic approach to generate sufficiently diverse feature maps and kernel instances, so that the evaluation of the metrics reveals the extent to which quantum mechanical characteristics are needed to solve the problem efficiently. Moreover, we need to define an effective and efficient process to perform the search for a (nearly) optimal quantum embedding, and this search process must be guided by the proxy metrics that we are trying to maximize.

The methodological workflow, illustrated in Fig. 1, formalizes the successive stages of this investigation, from dataset generation and quantum circuit parametrization to the evaluation of expressivity and entanglement metrics. This structured protocol ensures the controlled modulation of both data and model complexity, enabling reproducible experimentation and meaningful interpretation of observed effects. The results derived from this framework, presented in Sect. 4 constitute the foundation to

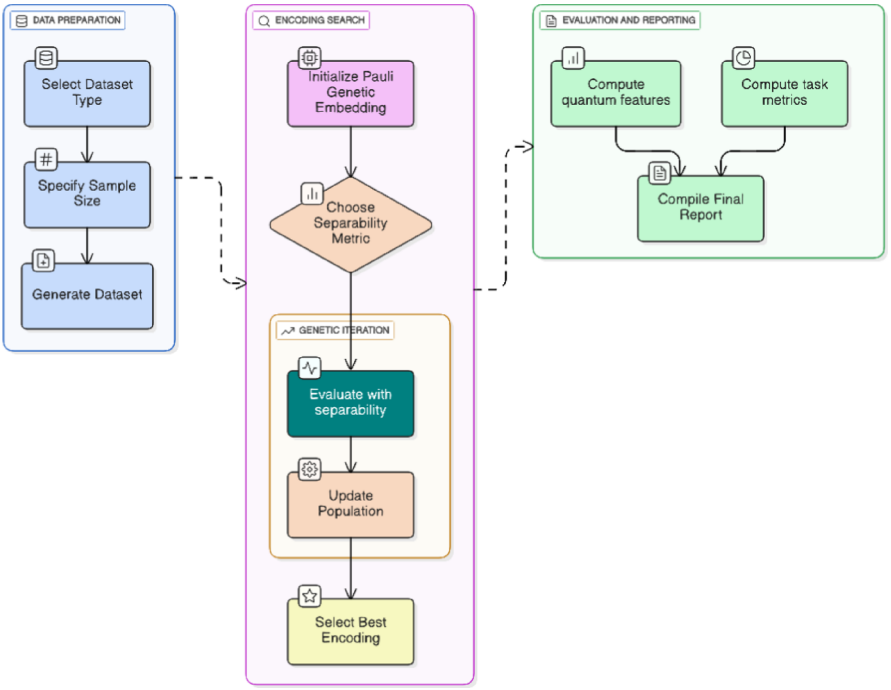


Fig. 1 Phases for each experiment: 1. dataset generation, 2. best feature map finding for a given proxy metric, and 3. task performance comparison

build strong evidence supporting the proposed quantitative relationship between data separability and the quantum mechanical characteristics of embedding circuits.

The current implementation allows for variations considering custom datasets or metric extensions for future works, making it composable and expandable with respect to the evolution of the field and knowledge gathered, but its end-to-end nature is described as follows:

1. *Data preparation*. This phase defines the input data used in the workflow. It involves selecting the dataset type, specifying the desired sample size, and generating the corresponding dataset that will serve as the basis for later encoding and evaluation.
   (a) *Select dataset type*: Choose the data distribution to be analyzed from available options.
   (b) *Specify dataset configuration*: Define the number of samples per class or overall dataset size, which can also control the level of class imbalance tunable by a ratio between classes.
   (c) *Generate dataset*: Create the dataset according to the specified parameters, ensuring reproducibility and consistency across experiments.

2. *Encoding search*. This phase focuses on identifying effective quantum embeddings through a genetic algorithm that evolves Pauli-based embeddings according to their separability performance.
   (a) *Initialize Pauli-Genetic embedding*: Generate an initial population of candidate embeddings represented as Pauli operator sequences.
   (b) *Choose separability metric*: Select a quantitative proxy metric to evaluate how well the embedding separates data classes in Hilbert space between KTA or CKA options.
   (c) *Genetic iteration*:
      (i) *Evaluate with separability*: Assess each embedding's fitness according to the chosen metric.
      (ii) *Update Population:* Apply genetic operators (mutation, crossover, selection) to evolve the population toward better-performing embeddings.
   (d) *Select best encoding*: Identify the optimal embedding configuration that will, expectedly, maximize separability on a tournament based strategy.
3. *Evaluation and reporting*. In the final phase, the selected encoding is analyzed in terms of its quantum properties and downstream task performance, with the results summarized in a comprehensive report benchmarking against strong classical candidates in the same regime.
   (a) *Compute quantum features*: Extract quantum feature representations generated by the selected embedding including quantum mechanical characteristics such as embedding entangling capacity or expressivity.
   (b) *Compute task metrics*: Evaluate the performance of these features on the target task (mainly through AUC of the classifier).
   (c) *Compile final report*: Summarize the findings, including best-performing embeddings and their metric values.

In the rest of this section, we discuss in detail how we have constructed suitable datasets and how we have designed the evolutionary search that we use to select adequate quantum embeddings. We also study the relationship between task complexity, circuit expressivity, and the sample size required for proper training.

All these details crystallize into the framework that can be used to systematically explore, evaluate, and compare quantum embeddings across different tasks.

3.1 Creating the datasets

To control the amount of complexity added to the dataset and incrementally check the contribution that a quantum embedding can provide, we opted for engineering the dataset to ensure full control over it. We build upon the proposal presented in [19] and we expanded it to a 3D version that allows us to explore states beyond the separable and maximally entangled regime that would be limited by the 2D nature of the original datasets. The datasets present 9 different situations with data points overlapping between different classes in an increasing complexity spectrum. From two simple blobs in the 1*a* category, periodically overlapping recurrent blobs in 2*c*, to blobs of one

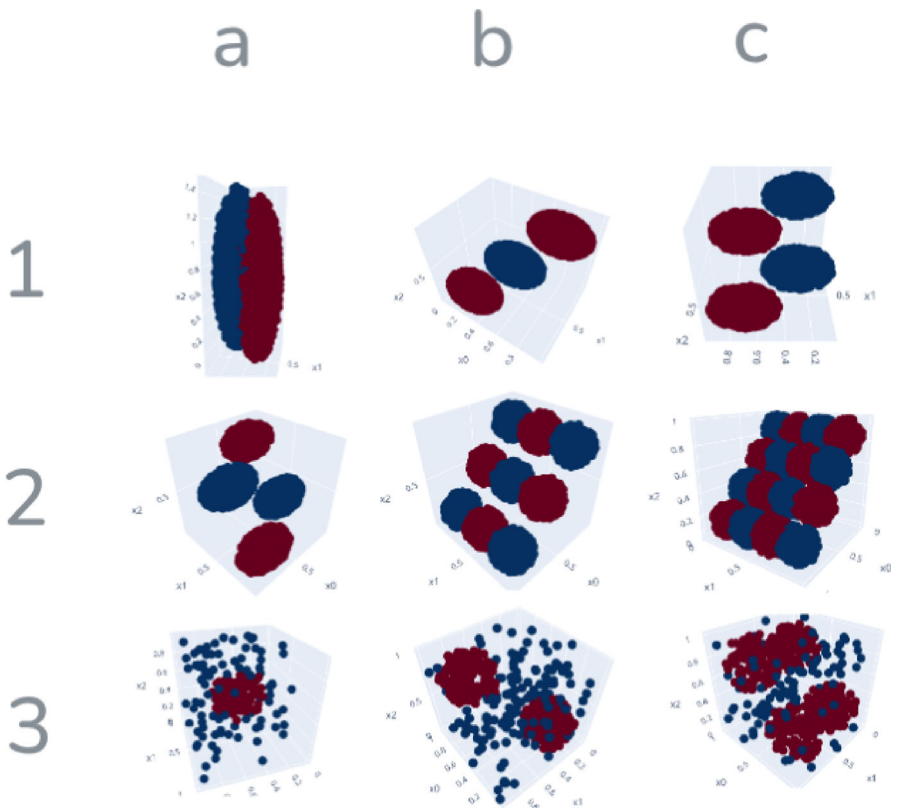


Fig. 2 Three-dimensional datasets increasing number of blobs and complexity by proximity of the blobs and embedding one class into another

category inside the cloud of points of the second category in datasets of type 3. These blobs can be modified in density of samples, proximity between blobs of different classes and radius for each of them.

The aim of manufacturing these datasets instead of using existing ones as raw entries is to be able to tweak the task complexity incrementally while avoiding any preprocessing that might remove nonlinear relationships existing among the features. Three-dimensional setups allow for the evaluation of multipartite entanglement and diverse mixtures of entangled states that are not present in two-body systems without incurring the complexity of scaling to higher-order datasets that may be harder to quantify in terms of complexity, either globally with respect to the task or locally with respect to the class overlap between blobs.

By modulating the proximity between the blobs and the density of each subclass, as shown in Fig. 2, we can tune the complexity required to create a hyperplane that separates both classes (depicted in blue and red in the figures). This ability to tweak and customize the task complexity is key to understanding how feature maps evolve during the search for an optimal embedding and the characteristics they present. In addition, by allowing control over the number of samples in each case, we can deliberately

adjust the degree of class imbalance—a characteristic that quantum machine learning is believed to address more effectively than classical methods.

While these datasets still lie within the regime of classically simulatable systems, they serve as a foundation for investigating the initial conditions that shape how particular data attributes may influence quantum embeddings, offering a pathway for systematic exploration in higher-dimensional feature space setups. The three-dimensional design of this initial dataset facilitates the study of multipartite relationships extending beyond maximally entangled states, offering an interpretable intermediary step before addressing higher-dimensional data spaces that require formal definitions of local complexity and periodic structures.

3.2 Evolutionary search through Pauli-Genetic Embedding

Once we have defined a way to create datasets with different complexities and relationship profiles as shown in Sect. 3.1, the next piece requires searching for optimal embeddings. Without any *a priori* knowledge, one needs to search the solution space trying to be as efficient as possible.

When exploring quantum embeddings through evolutionary algorithms, the absence of analytical gradients and the prohibitive cost of full model training for each candidate circuit require the definition of a proxy metric that can efficiently guide the search. Instead of computing classification accuracy for every possible embedding—an intractable process due to the exponential expansion of the quantum feature space—it is preferable to rely on an intrinsic measure that reflects how well a particular encoding structures the data in Hilbert space [11]. Such a proxy metric serves as a surrogate fitness function, enabling the evolutionary process to identify promising circuit configurations without solving the downstream learning task. To make this exploration feasible, each individual in the population must also be represented in a systematic and compact manner. Having a general enough formulation for each quantum embedding individually is key to fulfilling this requirement, encoding each embedding candidate as a composition of k-local Pauli interactions, thereby spanning all possible correlation orders among qubits while remaining interpretable and genetically manipulable. However, because the number of k-local interaction terms grows according to $\sum_{k=1}^{n} \binom{n}{k} = 2^n - 1$, the embedding space expands significantly with the number of qubits but also the number of layers or repetitions allowed for each embedding [2]. This combinatorial explosion renders exhaustive search infeasible even for moderate system sizes. Consequently, the evolutionary algorithm provides a principled stochastic mechanism to sample this high-dimensional landscape guided by the proxy metric, balancing exploration of new multi-qubit correlation patterns with exploitation of the most promising embeddings discovered throughout the search process.

Evolutionary search requires finding the right set of parameters not to perform a narrow solution space search, enabling both local optimization as well as global exploration. Looking into enabling this balance as per use case, our framework allows defining the search according to:

- Preprocessing stages for feature scaling or dimensionality reduction
- Maximum number of generations for the algorithm to run

- Population size for each generation
- Gene sequence identifying the feature map that allows for cross-breeding and mutation options
- Population selection based on tournament of maximally performing individuals at each generation

In order to enable manufacturing embeddings so that they can represent all sorts of classical-to-quantum state mapping strategies, a generic approach termed Pauli-Genetic Embedding (PGE) is proposed. It encompasses all possible k-qubit interactions, where k ranges from 1 to n for an n-qubit system encoding m samples with n-features each. The quantum circuit implementing this comprehensive embedding is expressed by

$$|\psi(X)\rangle = \left[\prod_{k=n}^{1} \prod_{S\in\binom{[n]}{k}} U_s\right] |0\rangle^{\otimes n} \tag{5}$$

where $\binom{[n]}{k}$ represents all possible k-element subsets of the set [n] = 1,2,...,n, $S = \{i_1, i_2, ..., i_k\}$ is a specific subset of k qubit indices, and U_S is the k-local unitary operation acting on qubits identified by indices in S. The k-local operations are defined by

$$U_S = \exp\left(-i \cdot \bigotimes_{j\in S} P_j \cdot \Theta_k(\{x_j\}_{j\in S})\right),$$

where $\bigotimes_{j\in S} P_j$ represents the tensor product of Pauli operators defined by the genetic encoding $\{I, X, Y, Z\}$ for an index instance in set S and $\Theta_k(\{x_j\}_{j\in S})$ is the parametrization function for k-local interactions.

The key to our generalized formulation is the parametrization function Θ_k that determines the rotation angle for k-local interactions. For example:

- Single-qubit operations (k=1):

$$\Theta_1(x_i) = 2(\pi - x_i)$$

- Two-qubit operations (k=2):

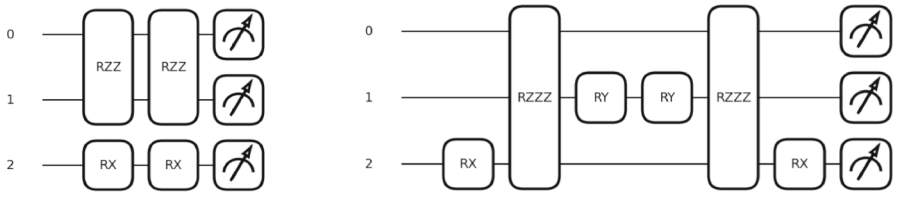
$$\Theta_2(x_i, x_j) = 2 \cdot (\pi - x_i) \cdot (\pi - x_j)$$

- Three-qubit operations (k=3):

$$\Theta_3(x_i, x_j, x_l) = 2 \cdot (\pi - x_i) \cdot (\pi - x_j) \cdot (\pi - x_l)$$

- k-qubit operations (general case):

$$\Theta_k(\{x_j\}_{j\in S}) = 2 \cdot \prod_{j\in S} (\pi - x_j)$$



(a) IIXZZI decomposition (b) IIXZZZIYI decomposition

Fig. 3 Fidelity kernels for each gene chain on a 3-qubit circuit splits. Both examples show how each sequence is transformed into different combinations of k-local operations

This parametrization ensures that single-qubit operations directly encode individual sample values, and higher-order operations encode the joint influence of multiple samples. The factor of 2 normalizes the rotation angles below the $[0, 2\pi)$ range, and the $(\pi - x_n)$ factor creates a distinction between low and high values in the interactions. These k-local interactions when aiming for hardware implementations would require further transpilation even though for the time being all operations are assumed to be run on a perfect device with the ability to render those k-local gates.

This generalized embedding exhibits several important mathematical characteristics:

- *Completeness*, by including all possible k-local interactions up to $k = n$, the embedding captures the complete correlation structure of the dataset.
- *Hierarchical information encoding* involves lower-order interactions that encode fundamental sample properties, while higher-order interactions capture complex relationships.
- *Parameter scaling*, given that the total number of parameters scales as $\sum_{k=1}^{n} \binom{n}{k} = 2^n - 1$, matching the exponential expressivity of the n-qubit Hilbert space.

The implementation is also relatively straightforward: for each gene sequence, one only needs to specify the corresponding Pauli chain that is then converted into the circuit. This representation allows cross-breeding and mutations required by genetic algorithms to be performed through simple string manipulations, enabling systematic exploration of all possible configurations in the worst-case scenario. New gene sequences can be set by cross-breeding the best-performing individuals, and mutations simply require a specific operator to change Pauli words to explore new embeddings. Thus, we can readily understand how data complexity affects the embedding in terms of the expressivity or capacity of entanglement requirements.

An example of the Pauli gates used on each qubit is shown in Fig. 3. Taking into consideration the example on Fig. 3a

3.3 Complexity and dataset size

The main motivation for creating a custom set of datasets is that the challenge they pose and the samples introduced can be tuned so that a fine-grained control over when quantum mechanical characteristics emerge depending on the characteristics of the

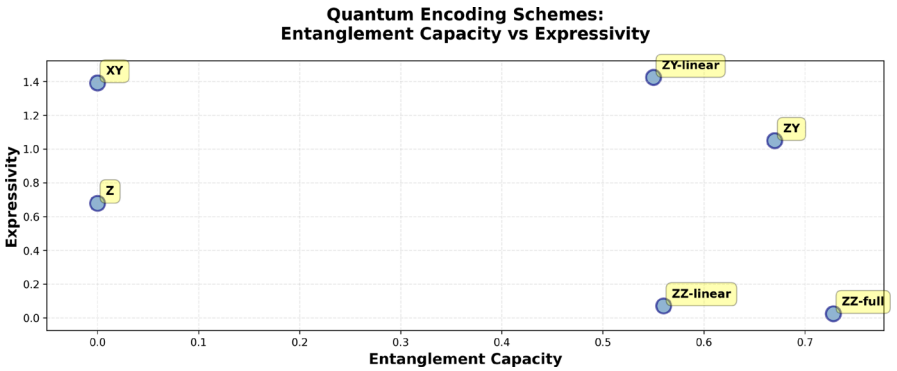


Fig. 4 Qiskit available encoding plotted according to their entanglement capacity and expressivity

challenge can be established, linking both quantum mechanical features to dataset complexity.

The PGE description of the embeddings, even though general enough, comes with some considerations that need to be highlighted for the proper use of the framework. It defines a function class $\mathcal{F}_{\text{PGE}}$ mapping input features $x_i \in [\frac{-\pi}{2}, \frac{\pi}{2}]^n$ to kernel evaluations $K(x_i, x_j) = |\langle\psi(x_i)||\psi(x_j)\rangle|^2$ whose complexity is bounded by Rademacher complexity theory principles [22]. For a PGE circuit with depth L, we establish that the Rademacher complexity of PGE circuits satisfies

$$\mathcal{R}_n(\mathcal{F}_{\text{PGE}}) \leq \sqrt{\frac{2L}{n}} \tag{6}$$

where $\mathcal{F}_{\text{PGE}}$ is the function describing the data encoding circuit obtained for each individual. This bound reveals a fundamental trade-off. The compositional nature of PGE circuits leads to Lipschitz constants that grow with depth. Deeper circuits increase expressivity but hurt generalization unless the training set size scales appropriately.

For a PGE chain whose sequence gets translated to $n \times L$, with L being the effective circuit depth, with k-local interactions we get

$$\text{Lip}(f_L) \leq L \cdot \max_k 2\pi^{k-1} \tag{7}$$

which implies a sample complexity requirement of

$$m \geq \mathcal{O}(L^2 \pi^{2(k_{\max}-1)}/\epsilon^2) \tag{8}$$

for generalization error ϵ. Thus, deeper circuits during the evolutionary search require exponentially more data. Thus, Rademacher complexity in worst-scenario cases of the search (Lipschitz constant $L \cdot \ell_k$) satisfies

$$\mathcal{R}_m(\mathcal{F}_{\text{PGE}}) \leq \frac{2L\pi^{k_{\max}-1}\sqrt{2^n}}{\sqrt{m}}. \tag{9}$$

The above-expressed PGE encoding defines an upper bound on the encoding layer described by the maximum chain size allowed during the evolutionary search, being the result of the gene length decomposition matching the number of available qubits to the effective L layers of independent k-local operations representing the circuit depth. Filling some of the terms with identity gates would decrease the number of layers, potentially being an empty operator if all gates are selected to be the Identity operation. But this is up to the evolution itself, requiring the maximum complexity to be matched with the dataset size m provided.

When it comes to generating custom datasets, width as well as the depth of the dataset needs to accommodate the defined upper ceiling learning requirements. This should also help understand some of the claims that establish QML as a better choice when facing scarcity of data, which may come from a combination of overfitted models and testing subset balance benefiting the summary metrics.

Through a systematic examination of circuit depth, expressivity, and entanglement capacity, we constructed a unified analytical framework designed to elucidate the bidirectional relationship between classical data complexity and the structural properties of quantum feature maps. This framework enables a principled exploration of how variations in circuit architecture and quantum resources—such as entanglement—affect the representational capacity of quantum embeddings when confronted with datasets of increasing intrinsic complexity. By quantitatively assessing these inter-dependencies, our approach establishes a methodological foundation for understanding when and how quantum feature maps can effectively capture nonlinear data relationships that challenge classical models due to the absence of quantum characteristics.

4 Experiments

4.1 Synthetic datasets

Given the cost of exploring the whole search space for embeddings allowing for all potential mappings into three-dimensional datasets (according to the PGE formulation in 5) three layer embeddings (to allow for independent gate contributions) were explored. Since the number of configurations grows exponentially as $N_{\text{PGE}} = 4^L$ for L layers, a heuristic approach to perform the search must be adopted. We aim to investigate feature maps that maximize the separability between samples in a classification task, without the need to train a complete model each time or to perform an exhaustive search over the entire set of possible feature maps. Hence the selection of a proxy metric such as CKA.

Considering Eq. 7 with a basic setup for 1- and 2-local terms which would be the gates to be used for a generalist hardware setup, we could fix $k_{\max=2}$. PGE encoding would allow for higher-order terms but in practice, all configuration will end translated into 1- and 2-qubit gates due to current hardware limitations. In a three-dimensional setup like the one in this work, where 3-local interactions are possible the 9-position gene sequence could cover three 3-local gate operations in the resulting feature map ($L = 3$). Thus, following Eq. 9 to achieve a complexity of $\mathcal{R}_m \ll 1$ approximately $m \geq 11,364$ samples would be needed. Not a feasible scenario for common computing

resources. In order to make the experiment trackable we challenge the framework to a poor generalization scenario of $m = 200$ which throws a complexity of $\mathcal{R}_{200} \approx \frac{53.3}{\sqrt{200}} = 3.77$ which would be in the memorization regime for any given model. This would also mean a good value in proxy metric during evolutionary search should fail when the classification metric is evaluated with test data.

Following the formulation introduced in Eq. 5, an evolutionary search was conducted independently over each dataset to identify quantum feature map configurations exhibiting the most favorable alignment between encoded data and target labels. This stochastic optimization strategy was designed to balance global exploration of the circuit search space with local refinement of promising candidates, avoiding the computational intractability of exhaustive enumeration while still covering a representative portion of the high-dimensional embedding landscape. The parameters used in this exploration were:

- No preprocessing done, just scaled features $\in [-\frac{\pi}{2}, \frac{\pi}{2}]$
- Maximum generations set to 100, with 10 iteration repetition of average performance for early stopping
- Population size of 100
- Pauli embedding capped to 3 layer depth feature maps
- Tournament selection on top 10 performant individuals
- Cross-breeding using 50% of the selected individuals
- Mutation probability of 5%

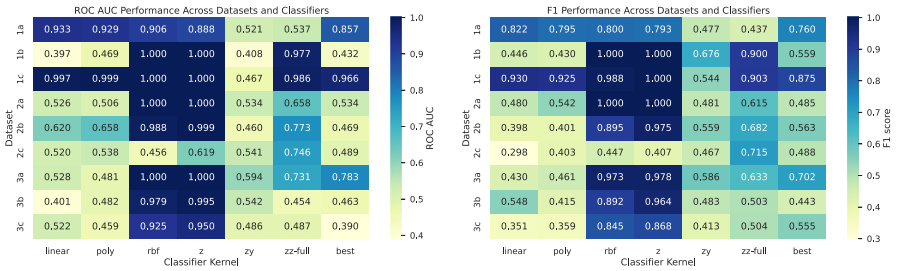
To contextualize the obtained results, the best-performing individuals from the evolutionary search were afterward benchmarked against standard classical and quantum baselines with some known fixed kernel options to compare against:

- Support Vector Classifiers (SVC) with linear, polynomial, and radial basis function kernels, using default hyperparameter configurations from the `scikit-learn` library [23].
- Quantum Support Vector Classifiers (QSVC) employing fixed non-optimized options implemented as single-repetition circuits, evaluated using precomputed kernel matrices available in the `Qiskit` framework [24].

The use of fixed embeddings in the latter quantum-classical benchmark ensured that the expressivity and entanglement capacity of the quantum feature maps remained dataset independent, allowing for direct comparison against the dynamically optimized embeddings generated by the evolutionary search even though layer structure may vary between evolved and fixed feature maps. Measuring the entanglement capacity and expressivity of each feature map shows a different profile in those axes as shown in Fig. 4.

Expressivity and entanglement capacity in Fig. 4 were discretely estimated to quantify the structural properties of the resulting quantum embeddings:

- *Expressivity*: evaluated using 120 bins, 4,000 randomly sampled parameter configurations, and 10,000 measurement shots.
- *Entanglement capacity*: computed over 4,000 samples per circuit.



(a) ROC AUC performance (b) F1 performance

Fig. 5 Metric heatmap per dataset and kernel type including the best found quantum kernel via evolutionary search and CKA proxy metric. Each cell represents the average metric value over 5 test sets in 80–20 (train-test) data splits

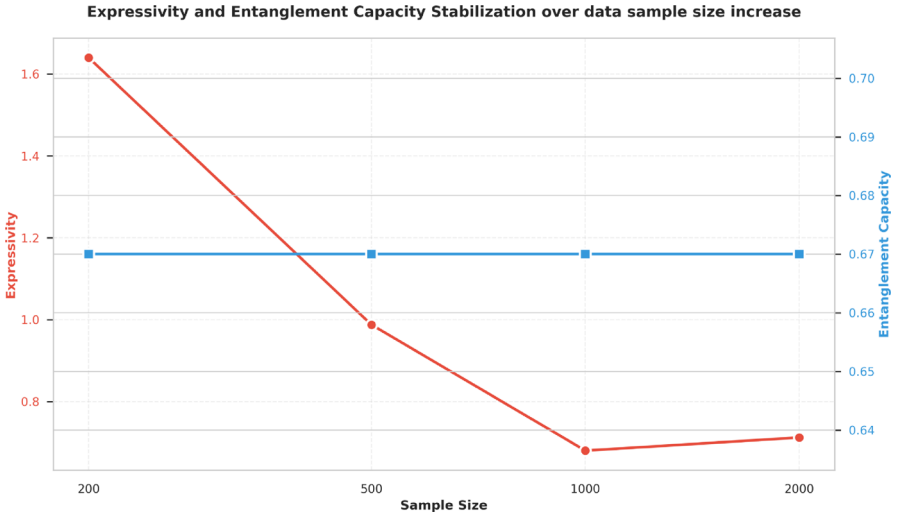
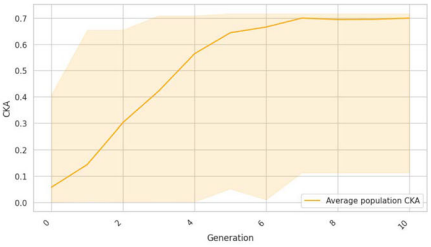


Fig. 6 Attribute stability for the 2*c* dataset scaled from 200 to 2000 samples after evolutionary search is performed extracting the average entanglement capacity and expressivity from top performing feature maps

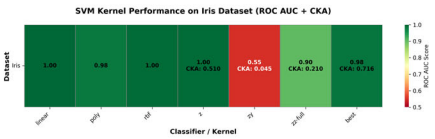
This configuration provides a practical compromise between computational feasibility and measurement precision, allowing the evolutionary process to explore meaningful differences in circuit behavior while avoiding the exponential cost of full state-space characterization.

A critical aspect when using evolutionary approaches is to guarantee enough capacity for locally optimizing the sampled space, but to allocate for exploratory processes looking for good spots far enough from the focal point. Comparing different kernels, including the best candidate obtained from the evolutionary search, Fig. 5 shows a comparison using classical and quantum kernels alongside Support Vector Classifier model average performance as the mean of 5 train–test data splits with different seeds.

Radial Basis Function (RBF) as well as Z kernel demonstrate their ability to encode diverse feature relationship while remaining generalist, but interestingly enough,



(a) Evolutionary search mean CKA evolution with minimum and maximum fitness shades.



(b) ROC AUC performance for various kernel choices averaged over 3 different seed train-test splits including averaged CKA for each quantum kernel type.

Fig. 7 Iris dataset evaluation focusing on separability between Virginica and Versicolor samples

dataset $2c$—that has a complex local pattern between data points of different classes—shows much better performance in the cases of ZZ feature map with full entanglement scheme ($L = 5$). The evolutionary search, focusing on higher CKA, did not correlate with better AUCs or F1 scores when considering different splits in this low-data scenario. Even though some low level correlation exists between proxy and performance metric, as pointed out by previous works, when tested against different splits the performance decays. This fact also allows us to highlight how easily overfitting may happen even for shallow quantum circuits due to their expressive nature.

Scaling up the analysis for the $2c$ case from the initial 200 samples up to 2000 samples where the Rademacher complexity starts reaching the point where generalization should occur, the CKA was maintained in a similar regime but the quantum characteristics seem to consolidate over a particular value, as seen in Fig. 6. Although larger sample sizes would need to be tested to confirm such hypothesis, which poses a computational effort worth revisiting in future works. The 2000 sample scenario with 100 individuals and 100 generations in the evolutionary setup already poses the challenge of simulating close to 400 billion independent circuits.

This periodic shape of the data seems to pose a complex enough pattern where entanglement and expressivity may offer a potential benefit over simpler approaches in the quantum regime or classical counterparts. It is worth further working on this case at higher scales as there might be an evident benefit coming not solely from the periodic nature of the data when it comes to generalization capabilities, but also the role of entanglement.

A closer examination of the feature-mapping circuits used to generate the kernels for the various datasets reveals that increasing circuit depth and gate count in pursuit of higher centered kernel alignment (CKA) does not necessarily translate into improved task performance. On the contrary, overly complex quantum embeddings tend to overfit the limited training data, capturing spurious correlations rather than task-relevant structures. This observation suggests that, within the regime of small or moderately sized datasets, highly expressive circuits may hinder rather than enhance generalization. Consequently, circuit parsimony maintaining a balanced trade-off between expressivity and model simplicity emerges as a more desirable design principle than unbounded growth in the number of quantum operations.

4.2 Iris dataset

Looking into datasets that might show similar characteristics but still fall in the range of simulatable experiments, we considered the Iris dataset, with 4 features and 50 samples per species. Particularly, we focused on the separability between Virginica and Versicolor labeled samples as a binary classification task. Larger datasets may not be suitable for the kind of study we are interested in, as one should either reduce the dimensionality potentially reducing the nonlinear relationships to linear mappings—as it often happens with PCA—or encode the data as it is, which for datasets like the Wine dataset (13 features) already needs a 13 qubit feature map requiring a depth exploration of several layers not to face the same Rademacher complexity boundaries. That is a task that would require distributed execution.

Given the non-exhaustive nature of the cases we have covered, no definitive conclusion can be derived from the relationship between expressivity, entanglement capacity and CKA. But it looks apparent from Fig. 7 that even though feature maps with entanglement capacity and expressivity may be used, a more simplistic approach may often lead to similar results in cases where memorization is the name of the game.

Interestingly enough, certain configurations, like the ZY, ended up being counterproductive despite their capacity of entanglement and expressivity surpasses the ones in simpler configurations (Z or ZZ mappings in particular).

Thus, in these particular situations, our results suggest that only a limited relationship between those characteristics exists in the scarce samples regime, providing a stronger case for periodical data embedding and overfitting scenarios than a quantum advantage via quantum characteristics. Therefore, a minimum sample size to data complexity ratio needs to be met. This seems to contradict the belief in the potential quantum machine learning advantage on scarce data scenarios, as it is equally likely that a pattern able to memorize the data has emerged.

5 Conclusions & future work

This work introduced the Pauli-Genetic Embedding (PGE) framework to systematically examine how quantum properties such as entanglement and expressivity influence learning performance in quantum machine learning (QML) classification tasks. Our empirical results show that quantum advantages in those scenarios are conditional rather than intrinsic—they depend strongly on data characteristics and sample complexity rather than on the presence of quantum correlations alone.

Despite achieving embeddings with high entanglement capacity (~ 0.72) and increased expressivity, quantum kernels generally matched or underperformed classical methods across our controlled three-dimensional datasets when it comes to generalization. The weak correspondence observed between quantum circuit metrics (e.g., CKA, expressivity, entanglement) during feature map search and classification accuracy on random splits indicates that current proxies may not capture task-relevant representational benefits or that sample sizes do actually matter for models aiming to generalize. Moreover, the evolutionary optimization process—while effective in

exploring the quantum feature space—frequently produced overfitted circuits, especially under data scarcity conditions. These findings challenge common claims that QML thrives under limited data regimes, suggesting instead that the apparent "quantum advantage" reported in such settings may stem from periodic or locally structured kernel behaviors rather than genuinely quantum representational power.

The proposed framework thus exposes a critical gap between theoretical quantum advantages and their empirical realization in practical learning scenarios. In particular, our results emphasize that intermediate alignment metrics such as CKA, though valuable for circuit search, do not reliably predict downstream generalization if minimum sample requirements are not met, underscoring the need for better measures linking quantum structure to learning efficacy and dataset characteristics. Additionally, the high expressivity of shallow circuits calls for explicit regularization and sample complexity-aware design to prevent spurious correlations from dominating optimization.

Future research should focus on identifying problem classes—such as our $2c$ instance—where quantum embeddings demonstrate consistent, empirical advantages. This requires both theoretical characterization of the conditions under which quantum feature maps can outperform classical kernels, and task-specific entanglement metrics that correlate directly with learning performance. Extending the framework to higher-dimensional feature spaces with adequate data and incorporating realistic noise and hardware topology models will be essential to assess near-term quantum feasibility.

Similarly, finding the way in which the framework would be able to scale up to cover datasets in the literature that may allow to drive stronger conclusions is required. Given the nature of the evolutionary search, naive task distribution can be performed but the scaling related to circuit simulation per process will need to be addressed within the field of distributable circuit simulation.

Finally, the field urgently needs standardized statistical evaluation protocols for QML, encompassing rigorous baselines, effect-size analysis, and reproducible benchmarks. Establishing these methodological foundations will enable the community to distinguish true quantum computational benefits from artifacts of data scarcity or circuit overfitting, moving QML from conceptual promise to empirically validated capability.

Acknowledgements This work has been partially supported by grant PID2023-146520OB-C22, funded by MICIU/AEI/10.13039/501100011033, by grant IDE/2024/000734 funded by Principado de Asturias, and by the Ministry for Digital Transformation and of Civil Service of the Spanish Government through the QUANTUM ENIA project call—Quantum Spain project; and by the European Union through the Recovery, Transformation and Resilience Plan—NextGenerationEU within the framework of the Digital Spain 2026 Agenda.

Author Contributions All authors contributed equally to this work. They were involved in the conceptualization, design, analysis, and preparation of the manuscript.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature.

Data Availability Data and code generated for these experiments can be found on https://github.com/IraitzM/kernel-trainer under version tag 0.1.8.

Declarations

Conflict of interest The authors have no Conflict of interest to declare that are relevant to the content of this article.

References

1. Biamonte J, Wittek P, Pancotti N, Rebentrost P, Wiebe N, Lloyd S (2017) Quantum machine learning. Nature. https://doi.org/10.1038/nature23474
2. Schuld M, Killoran N (2019) Quantum machine learning in feature hilbert spaces. Phys Rev Lett. https://doi.org/10.1103/physrevlett.122.040504
3. Schuld M (2021) Quantum machine learning models are kernel methods
4. Abbas A, Ambainis A, Augustino B, Bärtschi A, Buhrman H, Coffrin C, Cortiana G, Dunjko V, Egger DJ, Elmegreen BG, Franco N, Fratini F, Fuller B, Gacon J, Gonciulea C, Gribling S, Gupta S, Hadfield S, Heese R, Kircher G, Kleinert T, Koch T, Korpas G, Lenk S, Mareček J, Markov V, Mazzola G, Mensa S, Mohseni N, Nannicini G, O'Meara C, Tapia EP, Pokutta S, Proissl M, Rebentrost P, Sahin E, Symons BCB, Tornow S, Valls V, Woerner S, Wolf-Bauwens ML, Yard J, Yarkoni S, Zechiel D, Zhuk S, Zoufal C (2024) Challenges and opportunities in quantum optimization. Nat Rev Phys. https://doi.org/10.1038/s42254-024-00770-9
5. Mucke S, Heese R, Muller S, Wolter M, Piatkowski N (2022) Feature selection on quantum computers. Quantum Mach Intell. https://doi.org/10.1007/s42484-023-00099-z
6. Lubinski T, Coffrin C, McGeoch CC, Sathe P, Apanavicius J, Neira DEB (2023) Optimization applications as quantum performance benchmarks. ACM Trans Quantum Comput. https://doi.org/10.1145/3678184
7. Huang H-Y, Broughton M, Mohseni M, Babbush R, Boixo S, Neven H, McClean JR (2021) Power of data in quantum machine learning. Nat Commun. https://doi.org/10.1038/s41467-021-22539-9
8. Schnabel JA, Roth M (2025) Quantum kernel methods under scrutiny: a benchmarking study. Quantum Mach Intell. https://doi.org/10.1007/s42484-025-00273-5
9. Hofmann T, Schölkopf B, Smola AJ (2008) Kernel methods in machine learning. Ann Stat. https://doi.org/10.1214/009053607000000677
10. Rebentrost P, Mohseni M, Lloyd S (2014) Quantum support vector machine for big data classification. Phys Rev Lett. https://doi.org/10.1103/physrevlett.113.130503
11. Hubregtsen T, Wierichs D, Gil-Fuster E, Derks P-JHS, Faehrmann PK, Meyer JJ (2021) Training quantum embedding kernels on near-term quantum computers. Phys Rev A. https://doi.org/10.1103/physreva.106.042431
12. Incudini M, Bosco DL, Martini F, Grossi M, Serra G, Pierro AD (2024) Automatic and effective discovery of quantum kernels. IEEE Trans Emerg Topics Comput Intell. https://doi.org/10.1109/tetci.2024.3499993
13. Gujju Y, Harang R, Li C, Shibuya T, Zhao Q (2025) QuProFS: an evolutionary training-free approach to efficient quantum feature map search. arXiv:2508.07104
14. Pellow-Jarman R, Pillay A, Sinayskiy I, Petruccione F (2024) Hybrid genetic optimization for quantum feature map design. Quantum Mach Intell. https://doi.org/10.1007/s42484-024-00177-w
15. Phalak K, Ghosh A, Ghosh S (2024) Optimizing quantum embedding using genetic algorithm for qml applications. In: IEEE International Symposium on Quality Electronic Design. https://doi.org/10.1109/isqed65160.2025.11014359

16. Cortes C, Mohri M, Rostamizadeh A (2012) Algorithms for learning kernels based on centered alignment. J Mach Learn Res. https://doi.org/10.5555/2503308.2188413
17. Gühne O, Tóth G (2009) Entanglement detection. Phys Rep 474(1–6):1–75
18. Wei T-C, Goldbart PM (2003) Geometric measure of entanglement and applications to bipartite and multipartite quantum states. Phys Rev A. https://doi.org/10.1103/physreva.68.042307
19. Hubregtsen T, Pichlmeier J, Stecher P, Bertels K (2020) Evaluation of parameterized quantum circuits: on the relation between classification accuracy, expressibility, and entangling capability. Quantum Mach Intell. https://doi.org/10.1007/s42484-021-00038-w
20. Sim S, Johnson PD, Aspuru-Guzik A (2019) Expressibility and entangling capability of parameterized quantum circuits for hybrid quantum-classical algorithms. Adv Quantum Technol. https://doi.org/10.1002/qute.201900070
21. Brennen GK (2003) An observable measure of entanglement for pure states of multi-qubit systems. Quantum Inf Comput. https://doi.org/10.26421/qic3.6-5
22. Bu K, Koh DE, Li L, Luo Q, Zhang Y (2021) On the statistical complexity of quantum circuits. Phys Rev A. https://doi.org/10.1103/physreva.105.062431
23. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V et al (2011) Scikit-learn: machine learning in python. J Mach Learn Res 12:2825–2830
24. Javadi-Abhari A, Treinish M, Krsulich K, Wood CJ, Lishman J, Gacon J, Martiel S, Nation PD, Bishop L.S, Cross AW, Johnson BR, Gambetta JM (2024) Quantum computing with Qiskit. https://doi.org/10.48550/arXiv.2405.08810