

Provable advantages of kernel-based quantum learners and quantum preprocessing based on Grover's algorithm

T. Muser,^{1,*} E. Zapusek,^{1,*} V. Belis,² and F. Reiter¹

¹*Institute for Quantum Electronics, ETH Zürich, Otto-Stern-Weg 1, 8093 Zürich, Switzerland*

²*Institute for Particle Physics and Astrophysics, ETH Zürich, Otto-Stern-Weg 5, 8093 Zürich, Switzerland*



(Received 13 March 2024; accepted 12 June 2024; published 23 September 2024)

There is an ongoing effort to find quantum speedups for learning problems. Recently [Y. Liu *et al.*, *Nat. Phys.* **17**, 1013 (2021)] proved an exponential speedup for quantum support vector machines by leveraging the speedup of Shor's algorithm. We expand upon this result and identify a speedup utilizing Grover's algorithm in the kernel of a support vector machine. To show the practicality of the kernel structure we apply it to a problem related to pattern matching, providing a practical yet provable advantage. Moreover, we show that combining quantum computation in a preprocessing step with classical methods for classification further improves classifier performance.

DOI: [10.1103/PhysRevA.110.032434](https://doi.org/10.1103/PhysRevA.110.032434)

I. INTRODUCTION

Quantum computing has evolved from a theoretical concept to a physical reality, with experiments that demonstrate an advantage over classical computation [1–3]. The decades-long effort that led to these demonstrations is justified and driven by the performance guarantees of quantum algorithms such as Shor's and Grover's, which provide speedups for the problems of prime factorization and unstructured search [4,5].

The bustling subfield quantum machine learning promises to be one of the first applications of quantum computation. Leveraging algorithms with lower performance requirements, this domain presents exciting possibilities even with current quantum devices [6]. It is motivated by the tremendous success of classical machine learning, the potential power of quantum computers [7–9], and promising theoretical studies and guarantees [10–16]. However, for classical data the prospect of a quantum advantage with shallow quantum circuits is still unclear [17]. However, for some quantum machine learning algorithms which rely on fault-tolerant quantum computers a quantum speedup has been proven [18–22]. In particular, in Ref. [20], the authors Liu, Arunachalam, and Temme proved a quantum-classical separation for a quantum support vector machine (QSVM) by leveraging the speedup of Shor's factoring algorithm [20]. The proof is built on the widely held belief that there is no efficient classical algorithm for factorization. Just as Shor's and Grover's algorithms are vital for the field of quantum computing, algorithms with provable advantages are essential to the field of quantum machine learning as their quantum speedups pro-

vide theoretical justification for the effort invested in the field. They provide an example of a quantum advantage and thereby show promise for the existence of further quantum advantages [11,15,19,23].

With our present work, we follow a related approach and use Grover's algorithm to construct a QSVM. Grover's algorithm provides a quadratic speedup in query complexity for the problem of unstructured search. Discovered in 1996, it is one of the most prominent quantum algorithms, with applications in Boolean satisfiability [24–26], pattern matching, optimization [27–29], and quadratic unconstrained binary optimization problems [30]. Importantly, its speedup is provable [31], meaning that it can be rigorously shown that the best classical algorithm is quadratically slower than Grover's algorithm.

There is a strong connection between quantum machine learning and Grover's algorithm [32–38]. Many quantum machine learning algorithms use Grover's for the computationally expensive optimization during training. Models that can incorporate Grover's algorithm include support vector machines [32], perceptrons [35], clustering algorithms [39,40], and reinforcement learning [34].

In our study, we use Grover's algorithm in the kernel of a QSVM, allowing us to prove a quadratic speedup in a learning problem involving classical data [Fig. 1(a)]. We demonstrate the practical applications in Appendix A where we show that a kernel with the same structure can be applied to a learning problem of practical relevance; classifying texts or images.

Finally, we improve on the quantum kernel learner by instead supplying a classical model with data that has been preprocessed on a quantum machine. This allows intermediate states to be stored and removes the overhead required for noise-robustness thereby reducing the complexity. The same preprocessing approach can also be applied to improve the algorithm from Ref. [20]. In practical cases, one should use the classical learner in conjunction with quantum preprocessing, to achieve the lowest complexity.

*These authors contributed equally to this work.

[†]Present address: Swiss Data Science Center, ETH Zürich, Andreasstrasse 5, 8050 Zürich, Switzerland.

[‡]Contact author: zapuseke@ethz.ch

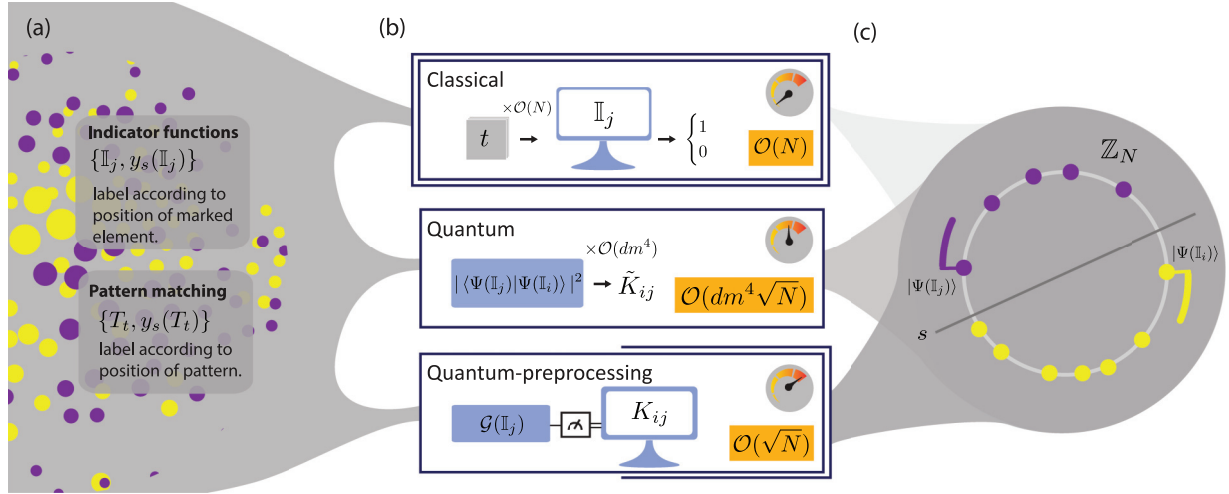


FIG. 1. (a) Description of the learning problem. We consider two distinct problems: Classifying black-box indicator functions based on the element they represent and classifying strings or images based on a pattern’s position. The approaches we apply are similar for both problems. Here we discuss the first problem in more detail. (b) Learning approaches. During inference the learner is presented with an unseen indicator function \mathbb{I}_j . A classical learner has to resort to brute-force sampling from the indicator function requiring $\mathcal{O}(N)$ queries. A quantum learner can use the quantum kernel, evaluating the inner product with d support vectors, thereby reducing the complexity to $\mathcal{O}(\sqrt{N})$. However, to ensure noise robustness one has to sample $\mathcal{O}(m^4)$ times, where m the size of the training set. Lastly, when the quantum processing is shifted to a preprocessing there is no noise robustness overhead. The output of the quantum circuit \mathcal{G} can be evaluated in a single run. Moreover, classical intermediate states can be stored and accessed later, eliminating the factor d . Consequently, the quantum preprocessing approach exhibits the lowest sample complexity. (c) Cyclic classification problem. The classification problem is a binary classification on the cyclic group \mathbb{Z}_N . The labels are assigned according to the value s .

II. QUANTUM SUPPORT VECTOR MACHINES

Support vector machines (SVMs) [41–43] are binary, linear classifiers. Provided with a set of training datapoints, each belonging to one of two classes, the SVM constructs a model that assigns class labels to unseen datapoints. For a given set of training examples an SVM deterministically finds a hyperplane that separates the datapoints with the maximal margin. To train an SVM, one needs to evaluate pairwise inner products of the training data $\mathbf{x}_i \in \mathcal{X}$. Classification of a new datum similarly requires the evaluation of inner products of the new datum with a subset of the training data, called *support vectors*. Nonlinear classification can be performed efficiently via the kernel trick [44,45]: We substitute the inner product with a function k , known as the *kernel*. This implicitly maps data with the *feature map* φ to an inner product of a different, high-dimensional Hilbert space \mathcal{H} , referred to as the latent or feature space

$$k(\mathbf{x}_i, \mathbf{x}_j) = \langle \varphi(\mathbf{x}_i), \varphi(\mathbf{x}_j) \rangle_{\mathcal{H}}. \quad (1)$$

A max-margin hyperplane will then be constructed in the feature space \mathcal{H} , corresponding to a potentially nonlinear decision boundary in the input data space \mathcal{X} . Using a kernel function k instead of an explicit feature map φ allows us to access spaces where calculating an explicit feature map would be expensive or even impossible. The space \mathcal{H} is hidden (or *latent*) and only accessed indirectly. In principle, a kernel allows classification of any kind of data, even data that are not numerical in nature (e.g., categorical data), as long as it can be mapped to some Hilbert space’s inner product.

In QSVMs the Hilbert space of some quantum system is used as the latent space. The feature map encodes a datum \mathbf{x}_i

into a quantum state $|\Psi(\mathbf{x}_i)\rangle$,

$$U(\mathbf{x}_i) : |0^n\rangle\langle 0^n| \mapsto |\Psi(\mathbf{x}_i)\rangle\langle \Psi(\mathbf{x}_i)|, \quad (2)$$

where $|0^n\rangle := |0\rangle^{\otimes n}$. The resulting state depends non-linearly on \mathbf{x}_i [8,23]. Now consider the circuit shown in Fig. 2. The probability K_{ij} of measuring the all-zero bitstring 0^n is equivalent to the Hilbert-Schmidt inner product

$$\begin{aligned} K_{ij} &= \text{Tr}[|\Psi(\mathbf{x}_i)\rangle\langle \Psi(\mathbf{x}_i)| \cdot |\Psi(\mathbf{x}_j)\rangle\langle \Psi(\mathbf{x}_j)|] \\ &= |\langle \Psi(\mathbf{x}_i) | \Psi(\mathbf{x}_j) \rangle|^2. \end{aligned} \quad (3)$$

It is an inner product in a Hilbert space \mathcal{H} and thus constitutes a valid kernel function

$$k(\mathbf{x}_i, \mathbf{x}_j) = K_{ij} = |\langle \Psi(\mathbf{x}_i) | \Psi(\mathbf{x}_j) \rangle|^2. \quad (4)$$

Since we cannot measure K_{ij} directly, we must instead estimate it through many measurements (or *shots*) R . This estimation is a stochastic process described by a binomial distribution and we find that finite-sampling bars us from evaluating the exact kernel. Instead, we only have access to

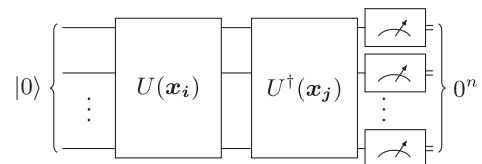


FIG. 2. Circuit realizing a quantum kernel. The probability of measuring the all-zero bitstring of length n , which is denoted by 0^n , is a similarity measure of $|\Psi(\mathbf{x}_i)\rangle$ and $|\Psi(\mathbf{x}_j)\rangle$.

a noisy version of the kernel

$$\tilde{K}_{ij} = K_{ij} + v_{ij}, \quad (5)$$

where v_{ij} describes the sampling noise with a variance

$$\text{Var}[v_{ij}] = \frac{K_{ij}[1 - K_{ij}]}{R} = O\left(\frac{1}{R}\right), \quad (6)$$

and mean zero.

III. GROVER'S ALGORITHM

Grover's algorithm [46] is one of the most well-known quantum algorithms due to its simplicity and versatility. Grover's Algorithm solves the problem known as *unstructured search*, where one needs to find one or multiple elements marked by a function f among a set with no discernible structure present. One can think of this situation arising, for example, when filtering a database for entries with a certain value for a specific key, with f being Boolean valued and 0 or false if the entry does not have the correct value and 1 or true if the value matches the search.

Classically, due to lack of structure, we can only evaluate f on one element at a time and it takes on average $\frac{N}{2}$ queries to find the marked element. As a result, the classical algorithm has query complexity of $O(N)$, scaling linearly with the size of the search space.

In the quantum case, we can, however, iteratively increase the amplitude of the marked states through an approach later generalized as amplitude amplification. This allows us to find the marked element with high probability in only $O(\sqrt{N})$ queries.

IV. CLASSIFICATION PROBLEM

The problem we want to solve revolves around the idea of classifying binary-valued *functions* based on which elements they assign to 0 and which to 1. Consider the space of the first $N = 2^n$ nonnegative integers \mathbb{Z}_N . We can define N different indicator functions $\mathbb{I}_j : \mathbb{Z}_N \rightarrow \{0, 1\}$ that mark a single element denoted by j ,

$$\mathbb{I}_j(t) = \begin{cases} 1 & \text{if } j = t, \\ 0 & \text{otherwise.} \end{cases} \quad (7)$$

For some $s \in \mathbb{Z}_{N/2}$ we assign classes to the functions \mathbb{I}_j based on j using the labeling rule or concept y_s ,

$$y_s(\mathbb{I}_j) = \begin{cases} +1 & \text{if } j \in [s, s + \frac{N-2}{2}], \\ -1 & \text{otherwise.} \end{cases} \quad (8)$$

Note that for $s \in \mathbb{Z}_N \setminus \mathbb{Z}_{N/2}$ the classes are equivalent to the classes of $s' = s - N/2$ with swapped class labels. The classification problem is described in Fig. 1(a) and visually represented in Fig. 1(c).

The learning problem can be summarized as follows: The quantum or classical learner will be provided with a training set of $m = \text{poly}(n)$ tuples $\{(\mathbb{I}_j, y_s(\mathbb{I}_j))\}$ consisting of functions and their corresponding label. After training has been completed the learner has to classify unknown \mathbb{I}_k .

The learning problem we investigate demonstrates a separation in the oracular setting of Grover's algorithm, i.e., the

training data are given as labeled black-box indicator functions. Towards practical applications, we discuss applying a related kernel to the real-world problem of pattern matching in Appendix A. This kernel, in turn, uses no oracles and input states are encoded as computational basis states. Although this approach makes weaker assumptions, we can still prove speedup in the worst-case complexity.

V. CONSTRUCTING A KERNEL

The classes contain functions that mark subsequent elements, therefore, we want to induce a bias into our kernel that assures that functions $\mathbb{I}_i, \mathbb{I}_j$ where $\|i - j\|$ is small have high kernel values. Consider to this effect a mapping to *interval states* [20]

$$U(\mathbb{I}_j) : |0\rangle \rightarrow |\Psi(\mathbb{I}_j)\rangle = \sqrt{\frac{1}{M}} \sum_{l=0}^{M-1} |j+l\rangle, \quad (9)$$

covering an interval of $2^m = M$ computational basis states. The construction of $U(\mathbb{I}_j)$ requires n qubits and n ancillas and can be split into four parts: First, we construct intervals in the first register by applying m Hadamard gates on the m least significant bits

$$H^{\otimes m} \otimes \mathbb{1} : |0^n\rangle \otimes |0^n\rangle \mapsto \left[\sqrt{\frac{1}{M}} \sum_l |l\rangle \right] \otimes |0^n\rangle. \quad (10)$$

Then, using Grover's algorithm [5] as a subroutine, we set the second register to the state corresponding to the marked element i . The success probability of Grover's algorithm exponentially converges to 1 in the size of the search space and is therefore assumed to be perfect in the following steps. We find, neglecting the normalization,

$$\mathbb{1} \otimes G_j : \sum_l |l\rangle \otimes |0^n\rangle \mapsto \sum_l |l\rangle \otimes |j\rangle. \quad (11)$$

By using a quantum adder Σ (see Appendix D) we add the second register to the first

$$\Sigma : \sum_l |l\rangle \otimes |j\rangle \mapsto \sum_l |j+l\rangle \otimes |j\rangle; \quad (12)$$

and subsequently uncompute the second register containing the ancillas using Grover's algorithm

$$\mathbb{1} \otimes G_j^\dagger : \sum_l |j+l\rangle \otimes |j\rangle \mapsto \sum_l |j+l\rangle \otimes |0^n\rangle. \quad (13)$$

It is vital that we are able to uncompute the ancillas, otherwise the adjoint of the feature map will not produce the appropriate state. This is possible as the algorithm outputs a classical state (cf. Appendix E). In Fig. 3 the quantum circuit of the feature map is shown. The quantum kernel (Fig. 2) implemented by this feature map is

$$k(\mathbb{I}_i, \mathbb{I}_j) = \begin{cases} \frac{M-|i-j|}{M} & \text{if } |i-j| \leq M, \\ 0 & \text{otherwise.} \end{cases} \quad (14)$$

However, we cannot measure the exact kernel value K_{ij} due to finite sampling and need to work with the noisy

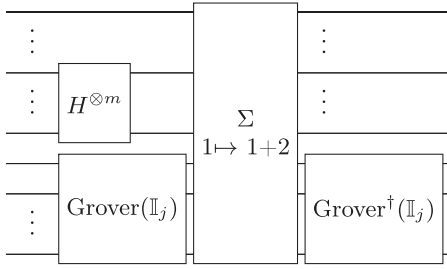


FIG. 3. The feature map $U(\mathbb{I}_j)$ in circuit form. The system is initialized in $|0^{2n}\rangle$. The upper register will contain the interval state $|\Psi(\mathbb{I}_j)\rangle$ that is used in the kernel, whereas the lower register is made up of ancilla qubits that can be discarded. A circuit for the quantum adder Σ is described in Appendix D.

estimate \tilde{K}_{ij} instead (cf. Sec. II). We can reduce the variance of \tilde{K}_{ij} by running the circuit R times since the variance v_{ij} is upper-bounded by $1/R$. Naturally, the question arises how many shots R are sufficient to ensure that the noise induced by finite sampling of the kernel does not strongly influence our classifier. We map our problem to the *noisy-halfspace learning problem* described in Ref. [20] to prove robustness against finite sampling noise (see Appendix B). In particular, this means that for training on $m = \text{poly}(n)$ data points, $R \sim O(m^4)$ shots are enough to bound the difference between the decision functions of classifiers trained with the exact kernel and noisy kernel below any desired constant with very high probability.

VI. CLASSICAL APPROACH AND QUERY COMPLEXITY

We will now compare the QSVM to a classical machine learning classifier by first deriving the bounds on the query complexity of a classical learner. We can easily construct a classical learner that uses a brute force approach to find the marked element of a given function. For a function $f : \mathbb{Z}_N \rightarrow \{0, 1\}$ marking a single element, this problem is solved using a linear search which takes, on average, $N/2$ guesses. A kernel such as Eq. (14) is easy to calculate classically once the element has been found and calculating a gram matrix for m samples generally necessitates $O(m^2)$ evaluations of the kernel. It follows that the query complexity of a classical kernel which sequentially searches for the marked element in the kernel is upper-bounded by $O(d\hat{l}N)$ during classification of \hat{l} new samples with d support vectors.

Alternatively, the unstructured search can also be performed as a preprocessing step, meaning that it is not part of the kernel. One can then use the processed data to evaluate the kernel according to Eq. (14) when needed. This reduces the complexity of classification to $O(\hat{l}N)$ since we already found the marked element for all support vectors d during training. As shown in Appendix C, any classical classifier solving this problem is lower-bounded by $O(N)$.

Now, let us consider the quantum SVM. The problem is then principally equivalent to the classical learner described above, with two major differences: First, we can identify the marked element using only $O(\sqrt{N})$ queries, however, we are working with an inexact, noisy kernel, requiring multiple

evaluations R . Combining these results, we find that it takes

$$O(d\hat{l}R\sqrt{N}) = O(d\hat{l}m^4\sqrt{N}) \quad (15)$$

evaluations to classify \hat{l} new samples. It is important to note that m scales with the problem size; in Appendix B, it is shown that $m = O[\text{poly}(n)]$.

If we have access to a quantum computer we can also consider a hybrid approach: During preprocessing we use Grover's algorithm to find the marked elements for each m training datapoints and \hat{l} unseen datapoints, save the results and use them to evaluate the classical kernel. This reduces the complexity further to $O(\hat{l}\sqrt{N})$. The preprocessing approach requires less computational resources since it only uses a single evaluation of Grover's algorithm compared to four for the fully quantum feature map described in Fig. 3.

In Fig. 1(b) the query complexity of the different approaches is shown for the classification of one new sample; $\hat{l} = 1$. The quantum approaches have a quadratic advantage over the classical learner. Limiting the quantum computation to a preprocessing step, rather than the fully quantum feature map, offers the best performance.

VII. CONCLUSION AND OUTLOOK

We identified a learning problem where quantum kernel methods offer a provable quadratic advantage compared to classical machine learning. Crucial to such advantages is biasing the kernel in a problem-specific way [47].

We arrived at our result using a feature map related to that described in Ref. [20], thereby demonstrating that their approach can be extended to other problems.

Grover's algorithm serves as an instructive example of this flexibility. Other algorithms can be used in place of Grover's algorithm. To demonstrate the practical use of our proposed strategy we apply our kernel in a learning problem involving pattern matching in Appendix A, achieving a quadratic speedup. Algorithms with a larger advantage could provide a larger speedup in the context of a learning problem [48–51].

Requiring the quantum subroutine to output a classical state, i.e., a basis state in a known basis, allows for the subsequent steps to be done classically. We show that the fully quantum kernel is inferior to a hybrid approach that utilizes a quantum computer for preprocessing, explicitly calculating the feature map, followed by a classical kernel calculation. This is in contrast with the original idea of the kernel trick where individual feature maps are hard (or impossible) to calculate directly, but the kernel function is computationally easy to evaluate.

Our findings provide an alternative perspective on the origin of quantum advantages in machine learning algorithms, specifically algorithms of the structure of Ref. [20]. If the quantum information processing can be shifted to a preprocessing step, thereby improving the performance of the algorithm, does it constitute a quantum advantage in learning [21]? We provide an explicit specification of an algorithm, using quantum preprocessing, that improves on the fully quantum learner. In practical problems, one is less concerned whether a performance advantage or computational speedup emerges from the learning model itself or the data

preprocessing steps. One can interpret the whole information processing pipeline, from the raw data preprocessing to the training of the statistical model, part of the learning algorithm. In real-world applications, the preprocessing steps, such as the normalization of the data, feature engineering and extraction [52,53], are crucial to the performance of the learner. Design choices made at the first steps of the pipeline can significantly hinder or enhance the performance of the learning model.

Along these lines, there is an increased focus on what is known as data-centric AI within the classical AI community [54–57]; for more details see Ref. [58] and references therein. This paradigm emphasizes the importance of the data preprocessing procedure in stark contrast to the conventional approach that focuses on the architecture of the model, size of the dataset, and the amount of computing power. There is much potential in the data-preprocessing steps: Only by adapting them in accordance with the structure of the data and the properties of the learner, the best performance can be achieved. Our work extends the ideas above by demonstrating, for the first time to our knowledge, the potential of quantum computing to aid in these crucial preprocessing steps. While the potential of quantum information processing for learning and for providing a speed-up in computationally expensive subroutines of classical machine learning algorithms has been established, our research highlights the benefits that quantum computing can offer as part of the data preprocessing procedure.

ACKNOWLEDGMENTS

We acknowledge discussion with C. Gyurik, V. Dunjko, and I. Rojko. F.R. and E.Z. acknowledge funding from the Swiss National Science Foundation (Ambizione Grant No. PZ00P2 186040) and the ETH Research Grant ETH-49 20-2. V.B. is supported by an ETH Research grant (Grant No. ETH C-04 21-2).

APPENDIX A: APPLICATION: PATTERN MATCHING

Pattern matching is a fundamental concept in computer science that refers to the process of searching for a specific pattern (i.e., a sequence of characters) within a larger piece of text. In the context of machine learning, pattern matching plays a critical role in tasks such as image and speech recognition and natural language processing [59–61].

The problem of pattern matching can be described as such: Given a text T of length N and a search string (pattern) P of length $L < N$ over a finite alphabet Σ . The task of the algorithm is to find an occurrence of P in T or to report that P is not a substring. We will discuss the problem in one dimension, however, it can be generalized to higher dimensions

The classical Knuth, Morris, and Pratt algorithm (Ref. [62]) solves the pattern matching problem in $O(L + N)$ worst-case time complexity. This is optimal, as to be correct for all inputs a classical algorithm has to check every character of pattern and text. For random strings Ref. [63] proves a classical lower bound of $\Omega(\frac{N}{L} \log_q L)$ where $|\Sigma| = q$, the size of the alphabet. The quantum algorithm presented in

Ref. [64] provides a solution of the problem with worst-case query complexity $O(\sqrt{N} + \sqrt{L})$ given oracular access to the text and pattern. For random patterns and texts quantum algorithms that are even faster on average are known [65].

In the construction of our kernel we focus on the algorithm from Ref. [66]. The data input are different from Ref. [64]: Pattern and text are provided as, classical, computational basis states rather than oracle queries. The algorithm worst-case time complexity $O\{\sqrt{N}[(\log(N))^2 + \log(L)]\}$. It requires $O(N + L)$ quantum space as it operates with classical data as an input to which it only has classical access. Here, we focus on optimizing time-complexity, so the space complexity is of no concern to us.

1. Description of learning problem

The learner is given a pattern P of length L and a set of texts $\{T_i\}$ of length N that are guaranteed to contain the search string. The index t denotes the (*a priori* unknown) location of the first bit of the pattern. We assign classes to the texts according to the position of the pattern P

$$y_s(T_j) = \begin{cases} +1 & \text{if } j \in [s, s + \frac{\log(N)-2}{2}], \\ -1 & \text{otherwise.} \end{cases} \quad (\text{A1})$$

The learner will be provided with a training set of $m = \text{poly}[\log(N)]$ tuples $\{(T_j, y_s(T_j))\}$ consisting of functions and their corresponding label. After training has been completed the learner has to classify an unseen text T_k . We summarize the learning problem in Fig. 1(a).

This application can be nicely motivated through tasks from image recognition. Instead of a one-dimensional string one can imagine a two-dimensional grid corresponding to pixels of an image. The classifier would then classify different images depending on where the searched pattern is located.

2. Summary of quantum algorithm for pattern matching

Here, we briefly summarize the algorithm presented in Ref. [66] for a more thorough analysis refer to the reference itself. We consider a tripartite quantum system of $n + N + L$ bits where $n = \log(N)$. The first register holds the index while second and third are used to store the text and pattern, respectively. The text T and the pattern P are encoded in the computational basis states of the second and third registers

$$|\mathcal{T}\rangle = |t_0 t_1 \dots t_{N-1}\rangle = \bigotimes_{i=0}^{N-1} |t_i\rangle, \quad (\text{A2})$$

$$|\mathcal{P}\rangle = |p_0 p_1 \dots p_{L-1}\rangle = \bigotimes_{l=0}^{L-1} |p_l\rangle. \quad (\text{A3})$$

For an alphabet of size 2 this can be done with identity and bitflip gates. We denote the encoding unitary as E . After encoding the state of the system is

$$|\psi_1\rangle = (\mathbb{1} \otimes E)|0^{n+N+L}\rangle = |0^n\rangle|\mathcal{T}\rangle|\mathcal{P}\rangle. \quad (\text{A4})$$

Assuming $N = 2^n$ we can apply a Walsh-Hadamard transform to the index register to generate a uniform superposition of

computational basis states

$$\begin{aligned} |\psi_2\rangle &= (H^{\otimes n} \otimes \mathbb{1})|\psi_1\rangle \\ &= \left(\frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} |k\rangle \right) \left(\bigotimes_{i=0}^{N-1} |t_i\rangle \right) \left(\bigotimes_{l=0}^{L-1} |p_l\rangle \right). \end{aligned} \quad (\text{A5})$$

Next, we apply the cyclic shift operator S that shifts the qubits of the text register by k sites according to the value of the index register $S|k\rangle \bigotimes_{i=0}^{N-1} |t_i\rangle = |k\rangle \bigotimes_{i=0}^{N-1} |t_{i+k}\rangle$. The implementation of the operator S is detailed in Ref. [66] and requires $O[\log(N)^2]$ time steps. The displaced state is

$$\begin{aligned} |\psi_3\rangle &= (S \otimes \mathbb{1})|\psi_2\rangle \\ &= \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} \left(|k\rangle \bigotimes_{i=0}^{N-1} |t_{(i+k)}\rangle \right) \left(\bigotimes_{l=0}^{L-1} |p_l\rangle \right). \end{aligned} \quad (\text{A6})$$

In the following step, a XOR operation is performed between the first L bits of the text register with the Pattern register. Such an operation can be performed unitarily using CNOT gates [67]. After this operation, we have

$$\begin{aligned} |\psi_4\rangle &= (\mathbb{1} \otimes \text{CNOT}^{\otimes L})|\psi_3\rangle \\ &= \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} \left(|k\rangle \bigotimes_{i=0}^{N-1} |t_{(i+k)}\rangle \bigotimes_{l=0}^{L-1} |p_l \oplus t_{(i+k)}\rangle \right). \end{aligned} \quad (\text{A7})$$

When the text that is displaced by the location of the pattern t it is mapped to the all zero state in the pattern register. The index register in turn is in the state t .

Finally, $O(\sqrt{N})$ steps of amplitude amplification are applied to the pattern register to amplify the amplitude of the $|0^L\rangle$ state. In total the algorithm requires $O\{\sqrt{N}[(\log(N))^2 + \log(L)]\}$ time steps. The final state is

$$|\psi_{\text{fin}}\rangle \approx |t\rangle \left(\bigotimes_{i=0}^{N-1} |t_{(i+t)}\rangle \right) (|0^L\rangle). \quad (\text{A8})$$

While overlap with states other than the desired output is $O(\frac{1}{N})$. We will denote the action of the pattern matching algorithm on input text T and pattern P as $\text{Pattern}(T, P)$.

3. Quantum kernel for pattern matching

The feature map $U_P(T_j)$ encodes a text into a superposition of indices that follow the matching index of a given pattern P in text T_j :

$$U_P(T_j) : |0\rangle \rightarrow |\Psi(T_j)\rangle = \sqrt{\frac{1}{M}} \sum_{l=0}^{M-1} |j+l\rangle. \quad (\text{A9})$$

To implement the feature map $U_P(T_j)$, a quantum circuit is used that includes initialization of the index superposition, a pattern matching subroutine, a quantum adder, and uncomputing of the pattern matching, as depicted in Fig. 4. This circuit is similar to the Grover feature map, but with the Grover subroutine replaced by pattern matching. The feature map described in Eq. (A9) implements a kernel

$$k(T_i, T_j) = \begin{cases} \frac{M-|i-j|}{M} & \text{if } |i-j| \leq M, \\ 0 & \text{otherwise.} \end{cases} \quad (\text{A10})$$

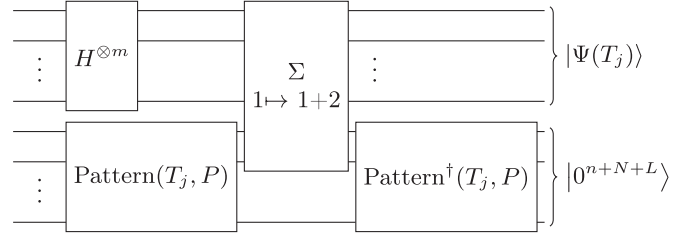


FIG. 4. The feature map $U_P(T_j)$ in circuit form. We divide the circuit into two registers of size n and $n + N + L$. The upper register will contain the interval state $|\Psi(T_j)\rangle$ that is used in the kernel, whereas the lower register is made up of ancilla qubits that can be discarded.

This kernel returns a value between 0 and 1, where 1 indicates a perfect match between the location of patterns in texts T_i and T_j , and 0 indicates a separation of more than M indices. The kernel calculation is performed on a quantum computer by applying the feature map $U_P(T_i)$ and $U_P^\dagger(T_j)$ to the initial state $|0^n\rangle$ and evaluating the probability of the all zero bitstring 0^n (Fig. 2).

In addition to text data, the feature map and kernel described in this work can be extended to other types of data, such as images. This is achieved by generalizing the pattern matching subroutine and interval states to multiple dimensions.

4. Classical hardness

The hardness of the problem depends on the input text. One can prove an advantage over a classical algorithm for pattern matching in the worst-case complexity. The worst-case complexity of classical pattern matching is $\Theta(N + L)$ [62]. Potentially, one has to check every character of the pattern and text. The quantum algorithm from Ref. [66] achieves the task in time $O\{\sqrt{N}[(\log(N))^2 + \log(L)]\}$. It thereby provides a quadratic speedup.

The quantum classical separation for the learning problem can be proven analogously to that in Appendix C. The hardness in classification depends on the texts and is therefore only valid in worst-case complexity.

APPENDIX B: SEPARABILITY AND THE NOISY-HALFSPACE LEARNING PROBLEM

We will show, analogously to Ref. [20], that the mapping Eq. (9) produces a separable problem. Define the *halfspace state* of s

$$|\Phi_s\rangle = \sqrt{\frac{2}{N}} \sum_{i=0}^{\frac{N}{2}-1} |s+i\rangle. \quad (\text{B1})$$

For ease of notation, we define the sets

$$c_s = \{\mathbb{I}_j | y_s(\mathbb{I}_j) = +1\}, \quad (\text{B2})$$

$$\bar{c}_s = \{\mathbb{I}_j | y_s(\mathbb{I}_j) = -1\}, \quad (\text{B3})$$

corresponding to the set of functions in one class or the other. Now consider a function f fulfilling the following conditions.

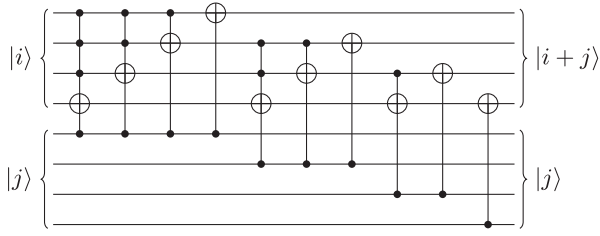


FIG. 5. Quantum adder circuit for two four-qubit registers. The lower register $|j\rangle$ is added to the upper register $|i\rangle$. Note that the uppermost qubit in each register corresponds to the lowest-order bit.

- (1) $f \in c_s$, i.e., $f(j) = \begin{cases} 1 & \text{if } j = t, \\ 0 & \text{otherwise.} \end{cases}$ for some $t \in [s, s + \frac{N-2}{2}]$
- (2) $s + \frac{N-2}{2} - t \geq M$, therefore all computational basis states of the superposition $|\Psi(f)\rangle$ are also present in $|\Phi_s\rangle$.

We define Δ as the overlap of the half-space state with $|\Psi(f)\rangle$;

$$\Delta = |\langle \Phi_s | \Psi(f) \rangle|^2 = \frac{1}{M} \left| \sum_{l=0}^{M-1} \underbrace{\langle \Phi_s | t+l \rangle \langle \Phi_s | t+l \rangle}_{=\sqrt{2/N}} \right|^2$$

$$= \frac{1}{M} \frac{2}{N} M^2 = \frac{2M}{N}. \quad (\text{B4})$$

It is easy to see that all but M functions in c_s also fulfill the second condition. In particular, for a fraction of

$$1 - \Delta = \frac{N/2 - M}{N/2} \quad (\text{B5})$$

the second condition and thus Eq. (B4) hold.

With the equivalent calculation for functions $f \in \bar{c}_s$ we find the following.

- (1) For a $1 - \Delta$ fraction of $f \in c_s$, $|\langle \Phi_s | \Psi(f) \rangle|^2 = \Delta$.
- (2) For a $1 - \Delta$ fraction of $f \in \bar{c}_s$, $|\langle \Phi_s | \Psi(f) \rangle|^2 = 0$.

By choosing $\mathbf{w}_s = |\Phi_s\rangle\langle\Phi_s|$ and $b = -\frac{\Delta}{2}$ we find a suitable hyperplane that separates the data in the latent space $\mathcal{H}(N)$. In particular, the probability of a data point lying on the wrong side of the margin or within it is Δ , the value of which we can essentially adjust freely by choice of a suitable M .

We see that this problem fulfills the conditions of the noisy half-space learning problem defined in [20]. From this it follows that for any s , training an SVM with the kernel k using $R \sim m^4$ shots will return a classifier with accuracy of 0.99 or greater on a test dataset with probability $2/3$ over random training samples and noise. Additionally, a number of $m = \text{poly}(n)$ training samples is sufficient for this purpose (see Ref. [20] Eq. (D40)).

APPENDIX C: CLASSICAL COMPLEXITY

Assume a classical learner is capable of determining the value s . During inference, the learner is provided with an unseen indicator function f . If the learner evaluates f on X values, corresponding to indicator functions of the same class, it finds the marked element with probability $\frac{X}{N}$. Knowing s , it can then assign the correct class. With probability $1 - \frac{X}{N}$ the correct marked element is not found, it then guesses the other class. This guess is correct with probability $P = \frac{1}{2} + \frac{X}{N}$. In total the classifier assigns the correct label with probability

$$P_s = \frac{X}{N} + \left(\frac{1}{2} + \frac{X}{N} \right) \left(1 - \frac{X}{N} \right) \leq \frac{1}{2} + \frac{3X}{2N}. \quad (\text{C1})$$

To classify f correctly with probability larger than some threshold (say 99%) we need that $X = O(N)$.

APPENDIX D: QUANTUM ADDER

The quantum adder Σ performs the action $\Sigma : |i\rangle \otimes |j\rangle \mapsto |i+j\rangle \otimes |j\rangle$ where $|i\rangle$ and $|j\rangle$ are computational basis states. A possible circuit that realizes such a unitary is pictured in Fig. 5 for four qubits.

APPENDIX E: ON THE EXTENSION TO INDICATOR FUNCTIONS THAT MARK MULTIPLE ELEMENTS

A natural extension to the Kernel and learning problem presented in the main text would be to extend it to indicator functions that mark multiple elements. However, for multiple reasons this extension fails.

The Kernel presented in Sec. V relies on the fact that after we apply Grover's algorithm the state is in or close to a single state in a known basis. Specifically, the computational basis. If one applies Grover to a function with multiple marked elements the output state is an even superposition of these elements. After applying the quantum adder (Fig. 5) the state of the ancilla system would be entangled with the first register. Due to this entanglement uncomputing the state on the ancilla register would also influence the main register. This would inhibit the proper function of the algorithm.

A different approach would be to not use the adder entirely. Provided that the functions mark sufficient elements the superpositions output by Grover's algorithm would have enough overlap to ensure noise robustness. However, the indicator functions would have to mark so many elements that classically solving the problem becomes easier, thereby removing any quantum advantage.

[1] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. S. L. Brandao, D. A. Buell *et al.*, Quantum supremacy using a programmable superconducting processor, *Nature (London)* **574**, 505 (2019).

[2] H.-S. Zhong, H. Wang, Y.-H. Deng, M.-C. Chen, L.-C. Peng, Y.-H. Luo, J. Qin, D. Wu, X. Ding, Y. Hu *et al.*, Quantum computational advantage using photons, *Science* **370**, 1460 (2020).

[3] L. S. Madsen, F. Laudenbach, M. F. Askarani, F. Rortais, T. Vincent, J. F. F. Bulmer, F. M. Miatto, L. Neuhaus, L. G. Helt,

- M. J. Collins *et al.*, Quantum computational advantage with a programmable photonic processor, *Nature (London)* **606**, 75 (2022).
- [4] P. W. Shor, Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer, *SIAM J. Comput.* **26**, 1484 (1997).
- [5] L. K. Grover, A fast quantum mechanical algorithm for database search, in *Proceedings of the Twenty-Eighth Annual ACM Symposium on Theory of Computing, STOC '96* (Association for Computing Machinery, New York, 1996), pp. 212–219.
- [6] A. J. Daley, I. Bloch, C. Kokail, S. Flannigan, N. Pearson, M. Troyer, and P. Zoller, Practical quantum advantage in quantum simulation, *Nature (London)* **607**, 667 (2022).
- [7] M. Benedetti, E. Lloyd, S. Sack, and M. Fiorentini, Parameterized quantum circuits as machine learning models, *Quantum Sci. Technol.* **4**, 043001 (2019).
- [8] M. Schuld and N. Killoran, Quantum machine learning in feature Hilbert spaces, *Phys. Rev. Lett.* **122**, 040504 (2019).
- [9] Y. Du, M.-H. Hsieh, T. Liu, S. You, and D. Tao, Learnability of quantum neural networks, *PRX Quantum* **2**, 040337 (2021).
- [10] A. Abbas, D. Sutter, C. Zoufal, A. Lucchi, A. Figalli, and S. Woerner, The power of quantum neural networks, *Nat. Comput. Sci.* **1**, 403 (2021).
- [11] H.-Y. Huang, M. Broughton, M. Mohseni, R. Babbush, S. Boixo, H. Neven, and J. R. McClean, Power of data in quantum machine learning, *Nat. Commun.* **12**, 2631 (2021).
- [12] T. Goto, Q. H. Tran, and K. Nakajima, Universal approximation property of quantum machine learning models in quantum-enhanced feature spaces, *Phys. Rev. Lett.* **127**, 090506 (2021).
- [13] M. C. Caro, H.-Y. Huang, M. Cerezo, K. Sharma, A. Sornborger, L. Cincio, and P. J. Coles, Generalization in quantum machine learning from few training data, *Nat. Commun.* **13**, 4919 (2022).
- [14] H.-Y. Huang, M. Broughton, J. Cotler, S. Chen, J. Li, M. Mohseni, H. Neven, R. Babbush, R. Kueng, J. Preskill, and J. R. McClean, Quantum advantage in learning from experiments, *Science* **376**, 1182 (2022).
- [15] H.-Y. Huang, R. Kueng, and J. Preskill, Information-theoretic bounds on quantum advantage in machine learning, *Phys. Rev. Lett.* **126**, 190505 (2021).
- [16] Y. Wu, B. Wu, J. Wang, and X. Yuan, Quantum phase recognition via quantum kernel methods, *Quantum* **7**, 981 (2023).
- [17] M. Cerezo, G. Verdon, H.-Y. Huang, L. Cincio, and P. J. Coles, Challenges and opportunities in quantum machine learning, *Nat. Comput. Sci.* **2**, 567 (2022).
- [18] R. A. Servedio and S. J. Gortler, Equivalences and separations between quantum and classical learnability, *SIAM J. Comput.* **33**, 1067 (2004).
- [19] P. Rebentrost, M. Mohseni, and S. Lloyd, Quantum support vector machine for big data classification, *Phys. Rev. Lett.* **113**, 130503 (2014).
- [20] Y. Liu, S. Arunachalam, and K. Temme, A rigorous and robust quantum speed-up in supervised machine learning, *Nat. Phys.* **17**, 1013 (2021).
- [21] C. Gyurik and V. Dunjko, On establishing learning separations between classical and quantum machine learning with classical data, [arxiv:2208.06339](https://arxiv.org/abs/2208.06339).
- [22] N. Pirnay, R. Sweke, J. Eisert, and J.-P. Seifert, Superpolynomial quantum-classical separation for density modeling, *Phys. Rev. A* **107**, 042416 (2023).
- [23] V. Havlíček, A. D. Córcoles, K. Temme, A. W. Harrow, A. Kandala, J. M. Chow, and J. M. Gambetta, Supervised learning with quantum-enhanced feature spaces, *Nature (London)* **567**, 209 (2019).
- [24] S. A. Cook, The complexity of theorem-proving procedures, in *Proceedings of the Third Annual ACM Symposium on Theory of Computing - STOC '71* (ACM, Shaker Heights, OH, 1971), pp. 151–158.
- [25] R. M. Karp, Reducibility among combinatorial problems, in *Complexity of Computer Computations*, edited by R. E. Miller, J. W. Thatcher, and J. D. Bohlinger (Springer, Boston, 1972), pp. 85–103.
- [26] W.-L. Yang, H. Wei, F. Zhou, W.-L. Chang, and M. Feng, Solution to the satisfiability problem using a complete Grover search with trapped ions, *J. Phys. B: At. Mol. Opt. Phys.* **42**, 145503 (2009).
- [27] C. Durr and P. Hoyer, A quantum algorithm for finding the minimum, [arXiv:quant-ph/9607014](https://arxiv.org/abs/quant-ph/9607014).
- [28] D. Bulger, W. P. Baritompa, and G. R. Wood, Implementing pure adaptive search with Grover's quantum algorithm, *J. Optim. Theory Appl.* **116**, 517 (2003).
- [29] W. P. Baritompa, D. W. Bulger, and G. R. Wood, Grover's quantum algorithm applied to global optimization, *SIAM J. Optim.* **15**, 1170 (2005).
- [30] A. Gilliam, S. Woerner, and C. Goniculea, Grover adaptive search for constrained polynomial binary optimization, *Quantum* **5**, 428 (2021).
- [31] C. H. Bennett, E. Bernstein, G. Brassard, and U. Vazirani, Strengths and weaknesses of quantum computing, *SIAM J. Comput.* **26**, 1510 (1997).
- [32] D. Anguita, S. Ridella, F. Riviello, and R. Zunino, Quantum optimization for training support vector machines, *Neural Networks* **16**, 763 (2003).
- [33] E. Aïmeur, G. Brassard, and S. Gambs, Quantum speed-up for unsupervised learning, *Mach. Learn.* **90**, 261 (2013).
- [34] G. D. Paparo, V. Dunjko, A. Makmal, M. A. Martin-Delgado, and H. J. Briegel, Quantum speedup for active learning agents, *Phys. Rev. X* **4**, 031002 (2014).
- [35] N. Wiebe, A. Kapoor, and K. M. Svore, Quantum algorithms for nearest-neighbor methods for supervised and unsupervised learning, *Quantum Inf. Comput.* **15**, 0318 (2015).
- [36] A. Kapoor, N. Wiebe, and K. Svore, Quantum perceptron models, in *Advances in Neural Information Processing Systems*, edited by D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett (Curran Associates, Red Hook, 2016), Vol. 29.
- [37] Y. Du, M.-H. Hsieh, T. Liu, and D. Tao, A Grover-search based quantum learning scheme for classification, *New J. Phys.* **23**, 023020 (2021).
- [38] W. Salmon, S. Strelchuk, and T. Gur, Provable advantage in quantum PAC learning, [arXiv:2309.10887](https://arxiv.org/abs/2309.10887).
- [39] S. U. Khan, A. J. Awan, and G. Vall-Llosera, K-means clustering on noisy intermediate scale quantum computers, [arXiv:1909.12183](https://arxiv.org/abs/1909.12183).
- [40] K. A. Woźniak, V. Belis, E. Puljak, P. Barkoutsos, G. Dissertori, M. Grossi, M. Pierini, F. Reiter, I. Tavernelli, and S. Vallecorsa,

- Quantum anomaly detection in the latent space of proton collision events at the LHC, [arXiv:2301.10780](https://arxiv.org/abs/2301.10780).
- [41] B. E. Boser, I. M. Guyon, and V. N. Vapnik, A training algorithm for optimal margin classifiers, in *Proceedings of the Fifth Annual Workshop on Computational Learning Theory* (ACM, New York, 1992).
- [42] C. Cortes and V. Vapnik, Support-vector networks, *Mach. Learn.* **20**, 273 (1995).
- [43] I. Steinwart and A. Christmann, *Support Vector Machines* (Springer, New York, 2008).
- [44] S. Bernhard and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond* (MIT Press, Cambridge, MA, 2009).
- [45] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*, 4th ed. (Academic, New York, 2009).
- [46] L. K. Grover, A fast quantum mechanical algorithm for database search, [arXiv:quant-ph/9605043](https://arxiv.org/abs/quant-ph/9605043).
- [47] J. Kübler, S. Buchholz, and B. Schölkopf, The inductive bias of quantum kernels, in *Advances in neural information processing systems*, edited by M. Ranzato, A. Beygelzimer, Y. Dauphin, P. S. Liang, and J. W. Vaughan (Curran Associates, Inc., New York, 2021), pp. 12661–12673.
- [48] A. M. Childs, R. Cleve, E. Deotto, E. Farhi, S. Gutmann, and D. A. Spielman, Exponential algorithmic speedup by a quantum walk, in *Proceedings of the Thirty-Fifth ACM Symposium on Theory of Computing - STOC '03* (ACM, San Diego, CA, 2003), p. 59.
- [49] G. Kuperberg, A subexponential-time quantum algorithm for the dihedral hidden subgroup problem, *SIAM J. Comput.* **35**, 170 (2005).
- [50] M. B. Hastings, A short path quantum algorithm for exact optimization, *Quantum* **2**, 78 (2018).
- [51] A. M. Dalzell, N. Pancotti, E. T. Campbell, and F. G. S. L. Brandão, Mind the gap: Achieving a super-Grover quantum speedup by jumping to the end, in *Proceedings of the 55th Annual ACM Symposium on Theory of Computing* (Association for Computing Machinery, New York, NY, USA, 2023).
- [52] J. Li, K. Cheng, S. Wang, F. Morstatter, R. P. Trevino, J. Tang, and H. Liu, Feature selection: A data perspective, *ACM Comput. Surv.* **50**, 1 (2017).
- [53] A. O. Salau and S. Jain, Feature extraction: A survey of the types, techniques, applications, in *2019 International Conference on Signal Processing and Communication (ICSC)* (ACM, New York, 2019), pp. 158–164.
- [54] M. Motamedi, N. Sakharykh, and T. Kaldewey, A data-centric approach for training deep neural networks with less data, [arXiv:2110.03613](https://arxiv.org/abs/2110.03613).
- [55] M. Hajjij, G. Zamzmi, K. N. Ramamurthy, and A. G. Saenz, Data-centric AI requires rethinking data notion, [arXiv:2110.02491](https://arxiv.org/abs/2110.02491).
- [56] L. Schmarje, V. Grossmann, C. Zelenka, S. Dippel, R. Kiko, M. Oszust, M. Pastell, J. Stracke, A. Valros, N. Volkmann, and R. Koch, Is one annotation enough? A data-centric image classification benchmark for noisy and ambiguous label estimation, [arXiv:2207.06214](https://arxiv.org/abs/2207.06214).
- [57] D. Zha, Z. P. Bhat, K.-H. Lai, F. Yang, Z. Jiang, S. Zhong, and X. Hu, Data-centric artificial intelligence: A survey, [arXiv:2303.10158](https://arxiv.org/abs/2303.10158).
- [58] NeurIPS Data-Centric AI Workshop, <https://datacentricai.org/neurips21/>.
- [59] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, G. Wang, J. Cai, and T. Chen, Recent advances in convolutional neural networks, *Pattern Recognit.* **77**, 354 (2018).
- [60] S. K. Saksamudre, P. Shrishrimal, and R. Deshmukh, A review on different approaches for speech recognition system, *Int. J. Comput. Appl.* **115**, 23 (2015).
- [61] K. R. Chowdhary, Natural language processing, in *Fundamentals of Artificial Intelligence* (Springer, New Delhi, 2020), pp. 603–649.
- [62] D. E. Knuth, J. H. Morris, Jr., and V. R. Pratt, Fast pattern matching in strings, *SIAM J. Comput.* **6**, 323 (1977).
- [63] A. C.-C. Yao, The complexity of pattern matching for a random string, *SIAM J. Comput.* **8**, 368 (1979).
- [64] H. Ramesh and V. Vinay, String matching in $o(\sqrt{n} + \sqrt{m})$ quantum time, [arXiv:quant-ph/0011049](https://arxiv.org/abs/quant-ph/0011049).
- [65] A. Montanaro, Quantum pattern matching fast on average, [arXiv:1408.1816](https://arxiv.org/abs/1408.1816).
- [66] P. Niroula and Y. Nam, A quantum algorithm for string matching, *npj Quantum Inf.* **7**, 37 (2021).
- [67] E. Zapusek, A. Javadi, and F. Reiter, Nonunitary gate operations by dissipation engineering, *Quantum Sci. Technol.* **8**, 015001 (2023).