Delta-Based Rare Pattern Discovery with Quantum Speedup

Simone Faro
Department of Mathematics and Computer Science
University of Catania
Catania, Italy
simone.faro@unict.it

Farida Farsian
Istituto Nazionale di Astrofisica
Catania, Italy
farida.farsian@inaf.it

Francesco Pio Marino
Department of Mathematics and Computer Science
University of Catania
Catania, Italy
francesco.marino@phd.unict.it

Gabriele Messina
Department of Mathematics and Computer Science
University of Catania
Catania, Italy
gabriele.messina@phd.unict.it

Francesco Schillirò
Istituto Nazionale di Astrofisica
Catania, Italy
francesco.schilliro@inaf.it

Eva Sciacca
Istituto Nazionale di Astrofisica
Catania, Italy
eva.sciacca@inaf.it

Fabio Vitello
Istituto Nazionale di Astrofisica
Catania, Italy
fabio.vitello@inaf.it

Abstract

Detecting rare and structurally meaningful patterns in time series is challenging, particularly when such patterns do not repeat exactly but exhibit variability due to noise or intrinsic dynamics. We introduce a novel framework for rare pattern discovery based on a combinatorial notion of approximate similarity over discretized signals. Our approach represents substrings through their local variations and defines pattern rarity in terms of the number of substrings that exhibit a similar evolution, using a tolerant Hamming distance over discrete differences. This delta-based representation captures structural behaviors such as bursts, trends, and oscillations, while providing robustness to noise and invariance to absolute signal levels, making it particularly suitable for real-world data such as astrophysical time series. We show that the resulting optimization problem is inherently quadratic in the classical setting, as it requires global pairwise comparisons between substrings and does not admit efficient incremental evaluation. To address this limitation, we propose a quantum approach that combines approximate counting and minimum finding, reducing the complexity from $O(n^2k)$ to $O(nk)$ and achieving a quadratic speedup in the sequence length. Our results identify a natural regime in which combinatorial pattern discovery aligns with the strengths of quantum algorithms, opening new directions for the analysis of complex temporal signals.

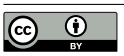
CCS Concepts

• **Theory of computation** → **Quantum computation theory**; *Quantum query complexity*; *Design and analysis of algorithms*; • **Information systems** → Data mining; • **Computing methodologies** → Pattern matching.

Keywords

Rare Pattern Discovery, Approximate String Matching, Quantum Algorithms, Time Series Analysis

ACM Reference Format:
Simone Faro, Farida Farsian, Francesco Pio Marino, Gabriele Messina, Francesco Schillirò, Eva Sciacca, and Fabio Vitello. 2026. Delta-Based Rare Pattern Discovery with Quantum Speedup. In *The 35th International Symposium on High-Performance Parallel and Distributed Computing (HPDC '26), July 13–16, 2026, Cleveland, OH, USA.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3806645.3816160



HPDC '26, Cleveland, OH, USA

ACM ISBN 979-8-4007-2640-8/26/07
https://doi.org/10.1145/3806645.3816160

1 Introduction

Astrophysical time series, such as those arising from Gamma Ray Burst (GRB) observations, exhibit complex temporal structures, including bursts, oscillations, and multi-peak behaviors [17]. GRBs are among the most energetic transient phenomena in the Universe, typically associated with the collapse of massive stars or compact object mergers, and are observed as short-lived flashes of gamma radiation followed by multi-wavelength afterglows [31]. Their temporal profiles are highly diverse, often characterized by irregular, multi-peaked structures with rapid variability, spanning timescales from milliseconds to several minutes, and exhibiting complex, non-stationary behavior. Identifying and characterizing such structures is essential for understanding the underlying physical mechanisms and for enabling data-driven discovery in high-energy astrophysics [31]. In many cases, the most informative events correspond to localized and irregular phenomena, which may occur rarely within otherwise regular or noisy signals [29].

Example 1.1. Consider a time series containing short bursts of activity with slightly different amplitudes and noise levels. For instance, the patterns $(2, 4, 5, 3)$, $(3, 5, 6, 4)$, and $(1, 3, 4, 2)$ exhibit a

similar rise-and-fall behavior, even though their absolute values differ.[1] Traditional exact or distance-based methods may fail to group these patterns together, or may treat them as distinct due to their amplitude differences. In contrast, by focusing on local variations, these patterns can be recognized as structurally similar. At the same time, a rare pattern corresponds to a behavior that appears only a few times under this notion of similarity, rather than to a simple outlier in terms of absolute values.

Traditional approaches to pattern detection in these signals include template matching, statistical thresholding, and, more recently, machine learning techniques such as neural networks. In the context of GRB analysis, machine learning approaches have been successfully applied to tasks such as burst detection, classification, and temporal pattern recognition, leveraging both light curves and sky maps to identify transient events and characterize their properties [28, 32]. In particular, deep learning architectures, including convolutional and recurrent neural networks, have demonstrated strong performance in capturing complex temporal structures in GRB signals, although they typically require labeled datasets and may suffer from limited interpretability [23, 24].

Template based methods rely on predefined shapes and are therefore limited in their ability to capture unexpected or novel behaviors. Statistical approaches, on the other hand, typically focus on global properties of the signal and may fail to detect subtle local structures. Learning-based methods, while powerful, require large amounts of labeled data, are sensitive to the choice of training distribution, and often lack interpretability, which is a critical requirement in scientific applications.

In parallel, the data mining community has developed a substantial body of work on rare pattern mining [2, 3], with the goal of identifying infrequent but potentially informative patterns in large datasets. Most existing approaches are based on frequency thresholds and combinatorial enumeration techniques, such as Apriori and FP-Growth. While effective in transactional settings, these methods are inherently designed for unordered data and do not directly extend to sequential or time-series domains. In particular, they do not account for temporal dependencies or for the local evolution of the signal, both of which are crucial in astrophysical contexts.

Time series data introduce additional challenges. First, relevant patterns are often not repeated exactly, but appear with variations due to noise, measurement uncertainty, or intrinsic physical variability. Second, the notion of similarity between patterns must take into account their temporal structure rather than their absolute values. Third, the search space of candidate patterns grows rapidly with the length of the signal, making exhaustive approaches computationally demanding.

In this work, we propose a novel framework for rare pattern discovery specifically tailored to time series data. Our approach departs from both classical frequency-based methods and learning-based models, and is based on three main components:

- a symbolic representation of time series,
- a delta-based characterization of local signal evolution,
- and a combinatorial formulation of approximate pattern frequency.

The key idea is to represent each substring through its sequence of local variations, rather than its absolute values, and to define similarity in terms of approximate matching between such representations. This leads to a notion of frequency based on the number of substrings that exhibit a similar local evolution up to a prescribed tolerance. As a result, the method is invariant to offsets, robust to noise, and capable of capturing meaningful structures such as bursts, trends, and oscillations.

From an algorithmic perspective, this formulation naturally leads to a quadratic search problem, as it requires comparing each substring with all others in the sequence. Moreover, due to the non-additive nature of the similarity measure, classical optimization techniques such as incremental updates or prefix-based pruning cannot be directly applied.

To address this intrinsic computational bottleneck, we explore the use of quantum algorithms. In particular, we combine approximate counting techniques with quantum minimum finding, enabling the identification of rare patterns with a quadratic speedup over classical exhaustive methods. This results in an overall complexity of $O(nk)$ compared to the classical $O(n^2k)$, where n is the length of the sequence and k is the pattern length.

Overall, our work provides a new perspective on rare pattern discovery in time series, bridging combinatorial pattern mining and quantum algorithms, and opening new possibilities for the analysis of complex astrophysical signals.

2 Related Work

Rare pattern mining has traditionally focused on identifying infrequent itemsets in transactional databases using algorithms such as Apriori, FP-Growth, and their extensions [1, 3, 11–13]. These approaches rely on support thresholds and combinatorial enumeration, often exploiting anti-monotonicity properties to prune the search space. While effective in structured, tabular settings, they are inherently designed for unordered data and exact matching, and do not naturally extend to sequential or time-series domains.

A well-known limitation of these methods is the so-called *rare item dilemma* [20], where lowering the support threshold leads to a rapid growth in the number of candidate patterns, making the mining process computationally expensive and often impractical. In addition, these approaches struggle with sparse datasets and do not account for situations in which patterns are not repeated exactly, but rather appear with variations. This limitation is particularly relevant in time-series data, where temporal dependencies and local signal dynamics play a fundamental role.

Time Series Pattern Discovery. In the context of time series analysis, pattern discovery is typically addressed through techniques such as motif discovery, shape-based matching, and symbolic representations [19, 30]. These methods generally rely on distance measures over raw or normalized signals [16] and are primarily designed to identify frequent or representative patterns. Approaches such as the Matrix Profile provide efficient mechanisms for motif and discord discovery, while symbolic representations like SAX enable scalable indexing and comparison over discretized signals.

[1] More precisely, similarity is evaluated through the differences between consecutive values rather than through the absolute values themselves. For instance, the sequences $(2, 5, 6, 4)$ and $(2, 4, 5, 3)$ are not considered similar, since their local variations differ: $5 - 2 \neq 4 - 2$.

Approach	Domain	Similarity	Variability	Rare	Training	Complexity
Apriori / FP-Growth	Transactional	Exact	No	Yes	No	Exponential
Motif / Matrix Profile	Time series	Distance-based	Yes	No	No	$O(n^2)$
SAX-based methods	Time series	Symbolic + distance	Yes	No	No	$O(n \log n)$
String matching	Sequences	Edit / Hamming	Yes	No	No	$O(nm)$
LSH / ANN	High-dim	Approx. metric	Yes	No	No	Subquadratic
Deep learning	Time series	Learned	Yes	No	Yes	High
Quantum ML	Time series	Learned	Yes	No	Yes	High
This work	Time series	Delta-based	Yes	Yes	No	$O(nk)$ (quantum)

Table 1: Comparison of existing approaches for pattern discovery and anomaly detection.

Similarly, shape-based matching methods based on dynamic time warping allow flexible alignment between subsequences [25].

Closely related are discord discovery methods, which aim at identifying anomalous subsequences that are maximally distant from all others [15]. While effective for anomaly detection, these methods rely on global distance criteria and do not explicitly capture structural similarity under local variations. As a result, existing time-series techniques are generally better suited to identifying frequent or highly distinctive patterns, rather than rare patterns exhibiting approximate structural similarity.

Combinatorial and String Processing Approaches. From a combinatorial and string processing perspective, a large body of work has addressed approximate pattern matching and similarity search in sequences. Classical techniques include dynamic programming approaches for edit distance and its variants, bit-parallel algorithms, and filtering-based methods for approximate matching [21, 22]. These methods are highly effective when the similarity measure is metric or decomposable, allowing incremental computation, pruning, or indexing. More advanced approaches exploit indexing structures, such as suffix trees, suffix arrays, or q-gram based filtering [10, 18], as well as locality-sensitive hashing (LSH) for approximate nearest neighbor search in high-dimensional spaces [14]. In time-series analysis, symbolic representations such as SAX enable the use of string matching techniques over discretized signals, often combined with indexing distances to accelerate search [19].

However, these techniques rely critically on structural properties that are not present in our setting. In particular, they typically assume that the similarity measure is either additive, metric, or admits a decomposition that allows partial evaluation and pruning [22]. In contrast, the delta-based similarity function d_τ used in this work is defined through position-wise thresholded comparisons and induces a non-metric, non-additive structure. As a consequence, standard filtering, indexing, and incremental update techniques cannot be directly applied.

Furthermore, the objective function $f(x_i)$ counts the number of substrings satisfying a global similarity constraint, which inherently requires comparing each candidate against all others. This prevents the use of classical acceleration strategies based on sliding windows, prefix reuse, or candidate pruning, and leads to a quadratic number of pairwise comparisons in the worst case. This intrinsic lack of structure is precisely what makes the problem particularly well suited to quantum algorithms for unstructured search and counting [4, 9].

Learning-Based Approaches. More recently, anomaly detection in time series has been approached using machine learning techniques, including neural networks and autoencoders [6, 27]. While these methods are capable of capturing complex temporal dependencies, they typically require training data, are sensitive to the choice of training distribution, and often lack interpretability. This can be a significant limitation in scientific applications, where understanding the structure of detected patterns is as important as detecting them.

Quantum Machine Learning Approaches. Recent works have also begun to explore the application of quantum machine learning techniques to astrophysical time series, including GRB analysis. In particular, hybrid quantum-classical models such as quantum convolutional neural networks and variational quantum circuits have been investigated for GRB classification tasks, demonstrating the potential of low-parameter quantum models to capture relevant features in both light curves and sky maps [8, 33]. In another work, [26] presented a comparative study between classical and quantum autoencoders for short-duration GRB signal detection. However, these approaches remain largely supervised and data-driven, requiring labeled datasets and training procedures, and therefore differ fundamentally from the combinatorial and training-free framework proposed in this work.

Positioning of This Work. In contrast to existing approaches, our method identifies the rarest patterns under an approximate notion of similarity based on local signal evolution. Rather than focusing on globally dissimilar subsequences, we target patterns that occur infrequently among structurally similar behaviors, as captured by the delta-based representation.

Our approach departs from both classical rare pattern mining and learning-based methods by adopting an approximate matching framework over local variations. This defines a notion of frequency based on structural similarity, making it suitable for noisy time series, while remaining training-free and interpretable.

From an algorithmic perspective, the problem is inherently quadratic, as each pattern must be compared with all others. We address this using a quantum-enhanced solution based on approximate counting and minimum finding [5], achieving a quadratic speedup.

To the best of our knowledge, this combination has not been previously explored.

Table 1 summarizes the main characteristics of existing approaches. Classical data mining relies on exact matching and does not extend to sequential data, while time-series methods capture variability but focus on frequent patterns. Combinatorial techniques require metric structure, and learning-based approaches require training and lack interpretability.

In contrast, our framework enables training-free rare pattern discovery under a non-metric similarity model, with a natural quantum speedup to $O(nk)$.

3 Delta-Based Approximate Similarity

A central challenge in time-series analysis is the identification of patterns that are not exactly repeated, but instead exhibit similar structural behavior under noise, variability, or measurement uncertainty. In many real-world signals, and in particular in astrophysical data such as GRB observations, relevant events are characterized more by their local evolution (e.g., sudden increases, oscillations, or decay phases) than by their absolute values.

This motivates the need for a representation that captures local dynamics while being robust to offsets and small perturbations. To this end, we adopt a delta-based representation of the signal, combined with an approximate notion of similarity between substrings.

Let $S = s_1 s_2 \dots s_n$ be a sequence over a finite alphabet Σ, obtained by discretizing a time series.

We consider all substrings of fixed length k:

$$x_i = S[i..i+k-1], \quad 1 \le i \le n-k+1.$$

Rather than comparing substrings directly based on their values, we associate to each substring x_i a sequence of discrete differences:

$$\Delta x_i = (\delta_{i,1}, \delta_{i,2}, \dots, \delta_{i,k-1}), \quad \text{where } \delta_{i,t} = s_{i+t} - s_{i+t-1}.$$

This transformation removes absolute offsets and emphasizes local variations in the signal. As a result, substrings that differ by a constant shift, or that exhibit similar trends with small perturbations, are mapped to similar delta representations.

To quantify similarity between substrings, we define a tolerant Hamming-like distance between their delta sequences. Given a tolerance parameter $\tau \ge 0$, we define:

$$d_\tau(x_i, x_j) = \sum_{t=1}^{k-1} \mathbf{1}\left(\left|\delta_{i,t} - \delta_{j,t}\right| > \tau\right).$$

This distance counts the number of positions where the local variations of the two substrings differ beyond the tolerance τ. In other words, small fluctuations are ignored, while significant deviations are penalized.

Two substrings are considered locally similar if:

$$d_\tau(x_i, x_j) \le d,$$

where d is a fixed mismatch threshold controlling the allowed number of discrepancies.

Based on this notion of similarity, we define the *approximate frequency* of a pattern x_i as:

$$f(x_i) = \left|\left\{ j \;\middle|\; d_\tau(x_i, x_j) \le d \right\}\right|.$$

This quantity measures how many substrings exhibit a similar local evolution to x_i, rather than counting exact repetitions. This distinction is crucial in time-series settings, where patterns are rarely repeated identically and often appear with variations due to noise or intrinsic dynamics.

The rare pattern discovery problem is then formulated as:

$$x^* = \arg \min_{1 \le i \le n-k+1} f(x_i),$$

that is, the substring with the smallest number of approximate occurrences. In case multiple substrings attain the same minimum frequency, any one of them may be returned. Equivalently, one may define a rarity score:

$$\text{score}(x_i) = \frac{1}{f(x_i)}.$$

The proposed model presents several desirable properties for the analysis of real-world signals. First, it is invariant to absolute signal levels, since the delta representation depends only on differences. Second, it is robust to noise, as the tolerance parameter τ allows small fluctuations to be disregarded. Third, it is sensitive to structural patterns such as bursts, oscillations, and trends, which are naturally encoded in the sequence of local variations. Finally, the use of approximate matching enables the detection of meaningful patterns that may exhibit variability rather than exact repetition.

These properties make the model particularly suitable for astrophysical time series. In GRB signals, for instance, relevant phenomena are typically characterized by rapid changes, peaks, and oscillatory behaviors, which are more effectively captured by differences than by absolute intensity values.

Example 3.1. Consider the sequence $S = (3, 5, 6, 4, 2, 3, 5)$, and let $k = 4$. We extract two substrings: $x_1 = (3, 5, 6, 4)$, $x_2 = (2, 3, 5, 4)$. Their delta representations are: $\Delta x_1 = (2, 1, -2)$, $\Delta x_2 = (1, 2, -1)$. Let the tolerance be $\tau = 1$. We compare the delta sequences position-wise:

$$|2-1| = 1 \le \tau, \quad |1-2| = 1 \le \tau, \quad |-2-(-1)| = 1 \le \tau.$$

Thus, $d_\tau(x_1, x_2) = 0$, and the two substrings are considered similar despite having different absolute values.

In contrast, a substring exhibiting a different structural behavior, such as a monotonic increase, would yield a larger distance and would not be counted in the approximate frequency. This illustrates how the delta-based representation captures local trends while remaining robust to offsets and small perturbations.

3.1 Classical Computational Complexity

The computation of $f(x_i)$ requires comparing the substring x_i with all other substrings:

$$f(x_i) = \sum_{j=1}^{n-k+1} \mathbf{1}\left(d_\tau(x_i, x_j) \le d\right).$$

Each comparison involves evaluating the distance $d_\tau(x_i, x_j)$, which requires $O(k)$ time. Since there are $O(n)$ substrings, computing all values $f(x_i)$ requires: $O(n^2 \cdot k)$ time in the classical setting.

In several related problems on sequences, more efficient solutions can be obtained by exploiting additional structure. For instance, exact pattern matching admits near-linear time algorithms based

on prefix functions or automata, while distance-based similarity search can benefit from indexing structures, hashing techniques, or locality-sensitive hashing when the distance function is metric and decomposable. Similarly, sliding window techniques allow incremental updates when the objective function can be expressed in an additive or factorizable form.

However, these approaches do not readily apply in our setting. First, the delta-based representation removes absolute information and encodes patterns through local variations, making it difficult to define effective indexing keys or hashing schemes that preserve the approximate similarity relation. Second, the tolerant distance d_τ depends on position-wise constraints over the entire substring and involves a thresholded comparison, which breaks metric properties and prevents the direct use of standard similarity search techniques. Third, the objective function $f(x_i)$ is inherently non-additive, as it counts the number of substrings satisfying a global constraint, and therefore does not support incremental updates under sliding window schemes.

While approximate or data-dependent heuristics may reduce the practical cost in specific scenarios, their effectiveness relies on assumptions about the distribution or regularity of the data. In the general case, no compact representation or local update rule is known that avoids evaluating a quadratic number of pairwise comparisons.

As a result, in the absence of strong assumptions on the input, the problem remains inherently quadratic in the worst case, requiring $O(n^2 \cdot k)$ time in the classical setting.

4 Quantum Solution

The delta-based formulation introduced in Section 3 leads to a computational problem that is inherently global: the frequency of each candidate substring depends on its similarity with all other substrings in the sequence. This lack of locality prevents the use of classical optimization techniques based on incremental updates or factorization, and results in a quadratic number of pairwise comparisons.

This type of structure is particularly well suited to quantum algorithms. Indeed, quantum search and counting techniques are known to provide quadratic speedups for problems that require evaluating a function over a large, unstructured set of candidates. In this section, we show how to leverage these techniques to accelerate rare pattern discovery under the proposed model.

For clarity, we summarize the overall quantum procedure in Algorithm 1. The algorithm combines two standard quantum primitives: amplitude estimation for approximating the frequency of a candidate pattern, and Dürr–Høyer minimum finding for selecting the rarest one.

4.1 Problem Formulation

Let S be a sequence of length n and let k be a fixed pattern length. For each index $i \in \{1, \ldots, n-k+1\}$, we consider the substring $x_i = S[i..i+k-1]$.

We define a Boolean similarity predicate

$$g(i,j) = \begin{cases} 1 & \text{if } d_\tau(x_i, x_j) \le d, \\ 0 & \text{otherwise,} \end{cases}$$

Algorithm: Quantum Rare Pattern Discovery

Input: sequence S of length n, pattern length k, tolerance τ, threshold d
Output: index i^* minimizing the approximate frequency
Construct oracle $O(i,j)$ returning 1 iff $d_\tau(x_i, x_j) \le d$
Function EstimateFrequency(i):
  Prepare uniform superposition over $j \in \{1, \ldots, n-k+1\}$
  Apply amplitude estimation using $O(i,j)$
  Return estimate $\widetilde{f}(x_i)$
Apply quantum minimum finding over $i \in \{1, \ldots, n-k+1\}$
  using EstimateFrequency(i) as evaluation procedure
Return $i^* = \arg\min_i \widetilde{f}(x_i)$

Figure 1: Quantum rare pattern discovery via amplitude estimation and minimum finding. For each candidate substring, the algorithm estimates its approximate frequency using a quantum oracle encoding the similarity relation, and applies Dürr–Høyer minimum finding to return the index of the rarest pattern.

and the corresponding approximate frequency

$$f(x_i) = \sum_{j=1}^{n-k+1} g(i,j),$$

which counts how many substrings exhibit a similar local evolution.

Equivalently, one may consider the normalized frequency

$$p_i = \frac{f(x_i)}{n-k+1},$$

which represents the probability that a randomly chosen substring is similar to x_i.

The rare pattern discovery problem is to identify

$$x^* = \arg\min_i f(x_i).$$

We note that $f(x_i)$ is inherently global, as it depends on comparisons with all other substrings, and does not admit an additive or decomposable structure.

4.2 Quantum Oracle

We define a quantum oracle that encodes the similarity relation between substrings. Given indices (i,j), the oracle O acts as

$$O\,|i,j\rangle\,|0\rangle = |i,j\rangle\,|g(i,j)\rangle\,,$$

where

$$g(i,j) = \begin{cases} 1 & \text{if } d_\tau(x_i, x_j) \le d, \\ 0 & \text{otherwise.} \end{cases}$$

The implementation of O follows a fully reversible computation pipeline. For a given pair (i,j), the oracle proceeds through the following stages:

(1) **Substring access.** The substrings $x_i = S[i..i+k-1]$ and $x_j = S[j..j+k-1]$ are coherently loaded into quantum registers. This can be achieved either through a QRAM model

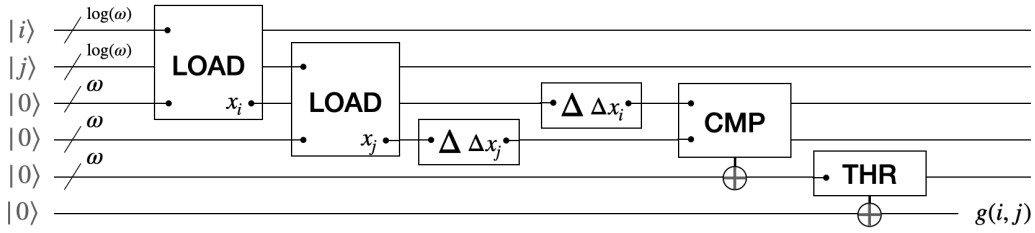


Figure 2: Quantum circuit implementing the oracle $O(i,j)$ for evaluating the similarity predicate $g(i,j)$. The circuit takes as input the indices i and j, coherently loads the corresponding substrings x_i and x_j, and computes their delta representations. It then evaluates the position-wise differences between the two sequences, counts the number of mismatches exceeding the tolerance τ, and compares the result against the threshold d. The output qubit encodes the value of $g(i,j)$, indicating whether the two substrings are considered similar. All components are realized using reversible arithmetic and comparison circuits.

or by assuming that the input sequence is stored in quantum memory, enabling parallel access in superposition.[2]

(2) **Delta computation.** The delta representations Δx_i and Δx_j are computed using reversible arithmetic circuits. Each difference $\delta_t = s_{i+t} - s_{i+t-1}$ is obtained via reversible subtraction, requiring $O(1)$ arithmetic operations per position. This step requires $O(k)$ elementary operations and $O(k)$ workspace qubits, which can be uncomputed after use.

(3) **Tolerance comparison.** For each position $t = 1, \ldots, k-1$, the circuit evaluates whether $|\delta_{i,t} - \delta_{j,t}| > \tau$ using reversible comparators. The result is stored in a single qubit indicating a mismatch at position t. This stage requires $O(k)$ comparisons and $O(k)$ auxiliary qubits.

(4) **Mismatch counting and thresholding.** The mismatch indicators are aggregated using a reversible adder to compute the total number of violations. The result is then compared against the threshold d using a reversible comparator, producing the output bit $g(i,j)$. This step requires $O(k)$ elementary gates and $O(\log k)$ ancilla qubits.

All intermediate computations can be uncomputed to restore ancillary registers to $|0\rangle$, ensuring that the oracle is implemented as a clean unitary transformation.

A schematic representation of the oracle $O(i,j)$ is shown in Figure 2. The circuit takes as input the indices i and j and coherently loads the corresponding substrings x_i and x_j. It then computes their delta representations, which encode the local variations of the signal, and evaluates the position-wise differences between the two sequences. The number of mismatches exceeding the tolerance τ is accumulated using a reversible counter, and the result is compared against the threshold d. The outcome of this comparison is stored in an output qubit, which encodes the value of $g(i,j)$ and indicates whether the two substrings are considered similar.

The modular structure of the circuit reflects the decomposition of the oracle into loading, transformation, and comparison stages. All components can be implemented using standard reversible arithmetic circuits, ensuring that the oracle can be evaluated coherently within superposition, as required by amplitude estimation.

Overall, the oracle can be realized with $O(k)$ arithmetic operations, $O(k)$ depth up to logarithmic factors depending on the adder implementation, and $O(k)$ auxiliary qubits. The dependence on the alphabet size is logarithmic in the bit-width of the encoded values.

We assume either a QRAM model or an equivalent input model in which the sequence is available in quantum registers. As is standard in the analysis of quantum search and counting algorithms, we focus on the cost of oracle invocations and internal arithmetic, abstracting from data-loading costs. We note that the oracle complexity remains linear in the pattern length k, and therefore does not affect the overall quadratic speedup in the sequence length n.

4.3 Frequency Estimation

For a fixed index i, computing $f(x_i)$ amounts to evaluating the number of indices j such that $g(i,j) = 1$, where $g(i,j)$ is the Boolean predicate implemented by the oracle O. Equivalently, $f(x_i)$ can be expressed as

$$f(x_i) = \sum_{j=1}^{n} g(i,j) = n \cdot \mathbb{E}_j[g(i,j)],$$

where the expectation is taken over the uniform distribution on $\{1, \ldots, n\}$.

We estimate this quantity using quantum amplitude estimation [5]. Specifically, we prepare the uniform superposition

$$\frac{1}{\sqrt{n}} \sum_{j=1}^{n} |j\rangle,$$

and use the oracle O to mark the "good" indices satisfying $g(i,j) = 1$. Amplitude estimation then provides an estimate $\tilde{p}_i$ of

$$p_i = \frac{f(x_i)}{n}$$

with additive error ε and constant success probability, using $O\left(\frac{\sqrt{n}}{\varepsilon}\right)$ oracle invocations.[3]

[2] As customary in quantum query complexity analyses, we abstract away the implementation cost of QRAM/data loading and assume coherent oracle access to the input sequence. The overall complexity therefore measures the number of oracle invocations and arithmetic operations, while the practical implementation cost may depend on the underlying memory architecture

[3] The parameter ε controls the precision of the amplitude estimation procedure. Smaller values of ε yield more accurate estimates of the approximate frequencies, at the cost of

Multiplying by n, we obtain an estimate $\tilde{f}(x_i)$ of $f(x_i)$ within additive error εn. Taking into account the $O(k)$ cost of each oracle evaluation, the overall complexity of frequency estimation is

$$O\left(\frac{\sqrt{n}}{\varepsilon} \cdot k\right).$$

The following lemma formalizes the complexity of the estimation procedure. Its proof follows directly from standard results on quantum amplitude estimation [5] and is therefore omitted for brevity. In particular, the estimation of $f(x_i)$ can be reduced to the problem of estimating the amplitude of the "good" states in a superposition over all candidate indices j, where a state is marked as good if it satisfies the similarity predicate encoded by the oracle. The additive error arises from the intrinsic precision of amplitude estimation, while the success probability can be amplified using standard repetition techniques.

Lemma 4.1. *For any fixed index i, the approximate frequency $f(x_i)$ can be estimated within additive error εn with constant success probability using $O\left(\frac{\sqrt{n}}{\varepsilon}\right)$ oracle queries.*

The above bound assumes a uniform superposition over all candidate indices and a single invocation of the estimation procedure. Higher success probabilities can be obtained by repeating the procedure a logarithmic number of times and taking the median of the estimates, without affecting the asymptotic query complexity up to constant factors.

4.4 Minimum Finding and Overall Complexity

To identify the minimum value of $f(x_i)$ over all candidate substrings, we employ the Dürr–Høyer quantum minimum finding algorithm [7]. Given access to an evaluation procedure for $f(x_i)$, this algorithm returns an index i^* minimizing $f(x_i)$ with constant success probability using $O(\sqrt{n})$ evaluations. Intuitively, the algorithm performs a quantum search over the set of candidate indices, progressively refining an upper bound on the minimum value through a sequence of comparisons, and exploiting amplitude amplification to achieve a quadratic speedup over classical exhaustive search.

In our setting, each evaluation of $f(x_i)$ is performed via amplitude estimation, as described in the previous section, with cost $O\left(\frac{\sqrt{n}}{\varepsilon} \cdot k\right)$. This reflects the need to estimate the number of indices j satisfying the similarity predicate, with precision εn, while accounting for the cost of evaluating the oracle. Combining the two procedures, we obtain an overall complexity

$$O(\sqrt{n}) \cdot O\left(\frac{\sqrt{n}}{\varepsilon} \cdot k\right) = O\left(\frac{n}{\varepsilon} \cdot k\right).$$

For constant precision ε, this yields a total complexity of $O(nk)$, matching the expected quadratic improvement over the classical baseline.

Lemma 4.2. *The rare pattern discovery problem can be solved with constant success probability using $O\left(\frac{n}{\varepsilon}\right)$ oracle calls and total time $O\left(\frac{n}{\varepsilon} \cdot k\right)$.*

a proportional increase in the number of oracle invocations. Throughout the analysis, we treat ε as a fixed constant, which is sufficient to distinguish the frequency gaps considered in our setting. However, instances involving extremely small or very close frequencies may require higher precision, potentially increasing the overall computational cost.

The above lemma establishes the main result of this work: rare pattern discovery under the proposed delta-based similarity model admits a quantum algorithm with overall time complexity $O(nk)$ for constant precision, achieving a quadratic speedup over the classical $O(n^2 \cdot k)$ approach.

The proof follows directly from the composition of amplitude estimation and quantum minimum finding [5, 7]. The overall success probability is obtained by combining the guarantees of the two procedures, and can be amplified through standard repetition techniques without affecting the asymptotic complexity.

In contrast, a classical exhaustive approach requires computing all pairwise similarities between substrings, resulting in a total complexity of $O(n^2 \cdot k)$. The proposed quantum algorithm therefore achieves a quadratic speedup in the sequence length n, which is optimal for unstructured search problems and consistent with known lower bounds.

4.5 Discussion

This speedup is not merely a consequence of applying a generic quantum search technique, but is tightly connected to the structure of the problem. The delta-based similarity measure induces a non-additive and non-factorizable objective function, which inherently requires global comparisons between substrings. While this limits the effectiveness of classical optimization strategies, it aligns naturally with the strengths of quantum algorithms for unstructured search and counting.

The quadratic speedup achieved by the proposed approach is consistent with known lower bounds for unstructured search problems. Since the objective function depends on global pairwise comparisons and lacks exploitable structure, no super-quadratic quantum speedup is expected in this setting. This further supports the natural alignment between the problem structure and quantum search techniques.

As a natural extension of this work, we plan to validate the proposed framework on both simulated and observational astrophysical datasets, with particular focus on GRB time series. These data provide a realistic testbed characterized by high variability, noise, and complex temporal structures, which are well aligned with the assumptions underlying the delta-based representation. In this context, we aim to assess the capability of the method to identify rare and physically meaningful patterns in real-world scenarios. In addition, a systematic comparison with existing approaches for anomaly and pattern detection will be conducted. This includes both classical techniques, such as statistical methods and machine learning-based models, and emerging quantum machine learning frameworks for anomaly detection.

5 Conclusion

We introduced a novel framework for rare pattern discovery in time series based on a delta-based representation and an approximate notion of similarity. By focusing on local signal variations rather than absolute values, the proposed model provides a natural way to capture structurally meaningful patterns in noisy and variable signals.

From a methodological perspective, the approach departs from both classical rare pattern mining and learning-based methods. Unlike frequency-based techniques, it allows for approximate matching and is therefore better suited to real-world time series where patterns are not repeated exactly. At the same time, it avoids the need for training data and yields explicitly interpretable results, which is particularly important in scientific applications.

From an algorithmic standpoint, we showed that the resulting formulation leads to a combinatorial optimization problem with intrinsic quadratic complexity. We then presented a quantum solution based on approximate counting and minimum finding, achieving a quadratic speedup with respect to the sequence length. This advantage is tightly connected to the non-additive and non-factorizable nature of the similarity measure, which limits the effectiveness of classical optimization techniques while aligning naturally with quantum search paradigms.

Overall, our work highlights a setting in which quantum algorithms can provide a concrete and meaningful advantage for data analysis tasks. The proposed framework opens several directions for future research, including the exploration of alternative similarity measures, the integration with more expressive symbolic representations, and the application to real astrophysical datasets.

More broadly, this work suggests that combining approximate pattern discovery with quantum algorithms may represent a promising direction for the analysis of complex temporal data.

References

[1] Rakesh Agrawal, Tomasz Imielinski, and Arun Swami. 1993. Mining Association Rules Between Sets of Items in Large Databases. In *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data*. 207–216. doi:10.1145/170035.170072

[2] Rakesh Agrawal, Tomasz Imielinski, and Arun N. Swami. 1993. Mining Association Rules between Sets of Items in Large Databases. In *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, Washington, DC, USA, May 26-28, 1993*, Peter Buneman and Sushil Jajodia (Eds.). ACM Press, 207–216. doi:10.1145/170035.170072

[3] Anindita Borah and Bhabesh Nath. 2019. Rare pattern mining: challenges and future perspectives. *Complex & Intelligent Systems* 5, 1 (2019), 1–23. doi:10.1007/s40747-018-0085-9

[4] Michel Boyer, Gilles Brassard, Peter Høyer, and Alain Tapp. 1998. Tight Bounds on Quantum Searching. *Fortschritte der Physik* 46, 4-5 (1998), 493–505. doi:10.1002/(SICI)1521-3978(199806)46:4/5<493::AID-PROP493>3.0.CO;2-P

[5] Gilles Brassard, Peter Høyer, Michele Mosca, and Alain Tapp. 2002. Quantum Amplitude Amplification and Estimation. *Contemp. Math.* 305 (2002), 53–74. doi:10.1090/conm/305/05215

[6] Raghavendra Chalapathy and Sanjay Chawla. 2019. Deep Learning for Anomaly Detection: A Survey. *arXiv preprint arXiv:1901.03407* (2019).

[7] Christoph Dürr and Peter Høyer. 1996. A Quantum Algorithm for Finding the Minimum. In *Proceedings of the 36th Annual Symposium on Foundations of Computer Science (FOCS)*. 454–462. doi:10.1109/SFCS.1996.548492

[8] Farida Farsian et al. 2025. Benchmarking Quantum Convolutional Neural Networks for Signal Classification in Simulated Gamma-Ray Burst Detection. arXiv:2501.17041 [astro-ph.HE] doi:10.1109/PDP66500.2025.00059

[9] Lov K. Grover. 1996. A Fast Quantum Mechanical Algorithm for Database Search. In *Proceedings of the 28th Annual ACM Symposium on Theory of Computing (STOC)*. 212–219. doi:10.1145/237814.237866

[10] Dan Gusfield. 1997. *Algorithms on Strings, Trees, and Sequences*. Cambridge University Press. doi:10.1017/CBO9780511574931

[11] Jiawei Han, Jian Pei, and Yiwen Yin. 2000. Mining Frequent Patterns without Candidate Generation. In *Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data, May 16-18, 2000, Dallas, Texas, USA*, Weidong Chen, Jeffrey F. Naughton, and Philip A. Bernstein (Eds.). ACM, 1–12. doi:10.1145/342009.335372

[12] Jiawei Han, Jian Pei, and Yiwen Yin. 2000. Mining Frequent Patterns without Candidate Generation. In *Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data*. 1–12. doi:10.1145/342009.335372

[13] Ya-Han Hu and Yen-Liang Chen. 2006. Mining association rules with multiple minimum supports: a new mining algorithm and a support tuning mechanism. *Decision Support Systems* 42, 1 (2006), 1–24. doi:10.1016/j.dss.2004.09.007

[14] Piotr Indyk and Rajeev Motwani. 1998. Approximate Nearest Neighbors: Towards Removing the Curse of Dimensionality. In *Proceedings of the 30th Annual ACM Symposium on Theory of Computing (STOC)*. 604–613. doi:10.1145/276698.276876

[15] Eamonn Keogh, Jessica Lin, and Ada Fu. 2005. HOT SAX: Efficiently Finding the Most Unusual Time Series Subsequence. In *Proceedings of the Fifth IEEE International Conference on Data Mining*. 226–233. doi:10.1109/ICDM.2005.79

[16] Eamonn Keogh and Chotirat Ann Ratanamahatana. 2005. Exact indexing of dynamic time warping. *Knowledge and Information Systems* 7, 3 (2005), 358–386. doi:10.1007/s10115-004-0154-9

[17] Pawan Kumar and Bing Zhang. 2014. The physics of gamma-ray bursts \& relativistic jets. *Phys. Rept.* 561 (2014), 1–109. arXiv:1410.0679 [astro-ph.HE] doi:10.1016/j.physrep.2014.09.008

[18] Stefan Kurtz. 1998. Reducing the Space Requirement of Suffix Trees. In *Proceedings of the 9th Annual Symposium on Combinatorial Pattern Matching (CPM)*. 111–122. doi:10.1007/BFb0030784

[19] Jessica Lin, Eamonn Keogh, Li Wei, and Stefano Lonardi. 2007. Experiencing SAX: a novel symbolic representation of time series. *Data Mining and Knowledge Discovery* 15, 2 (2007), 107–144. doi:10.1007/s10618-007-0064-z

[20] Bing Liu, Wynne Hsu, and Yiming Ma. 1999. Mining Association Rules with Multiple Minimum Supports. In *Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 337–341. doi:10.1145/312129.312277

[21] Gene Myers. 1999. A Fast Bit-Vector Algorithm for Approximate String Matching Based on Dynamic Programming. *J. ACM* 46, 3 (1999), 395–415. doi:10.1145/316542.316550

[22] Gonzalo Navarro. 2001. A Guided Tour to Approximate String Matching. *Comput. Surveys* 33, 1 (2001), 31–88. doi:10.1145/375360.375365

[23] N. Parmiggiani, A. Bulgarelli, V. Fioretti, A. Di Piano, A. Giuliani, F. Longo, F. Verrecchia, M. Tavani, D. Beneventano, and A. Macaluso. 2021. A Deep Learning Method for AGILE-GRID Gamma-Ray Burst Detection. *The Astrophysical Journal* 914, 1 (2021), 67. doi:10.3847/1538-4357/abf3c1

[24] N. Parmiggiani, A. Bulgarelli, A. Ursi, M. Tavani, A. Macaluso, A. Di Piano, V. Fioretti, L. Baroncelli, A. Addis, and C. Pittori. 2023. A Deep-learning Anomaly-detection Method to Identify Gamma-Ray Bursts in the Ratemeters of the AGILE Anticoincidence System. *The Astrophysical Journal* 945, 1 (2023), 34. doi:10.3847/1538-4357/acb7e4

[25] Thanawin Rakthanmanon, Bilson Campana, Abdullah Mueen, Gustavo Batista, Brandon Westover, Qiang Zhu, Jesin Zakaria, and Eamonn Keogh. 2012. Searching and Mining Trillions of Time Series Subsequences under Dynamic Time Warping. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 262–270. doi:10.1145/2339530.2339576

[26] Alessandro Rizzo, Nicolò Parmiggiani, Andrea Bulgarelli, Antonio Macaluso, Carlo Burigana, Eric Burns, Luca Cappelli, Vincenzo Fabrizio Cardone, Farida Farsian, Savitri Gallego, Giuseppe Murante, Giuseppe Sarracino, Roberto Scaramella, Francesco Schillirò, Vincenzo Testa, Tiziana Trombetti, Dieter H. Hartmann, Carolyn A. Kierans, Eliza Neights, John A. Tomsick, and Andreas Zoglauer. 2026. Comparing Classical and Quantum Deep Learning Techniques for Anomaly Detection of Short-Duration Gamma-Ray Signals. *Astronomy and Computing* 56 (2026), 101090. doi:10.1016/j.ascom.2026.101090

[27] Lukas Ruff, Robert Vandermeulen, Nico Görnitz, Alexander Binder, Emmanuel Müller, and Marius Kloft. 2021. Deep One-Class Classification. *International Journal of Computer Vision* 129 (2021), 1–23. doi:10.1007/s11263-020-01325-4

[28] L. Salmon, L. Hanlon, and A. Martin-Carrillo. 2022. Two Classes of Gamma-ray Bursts Distinguished within the First Second of Their Prompt Emission. *Galaxies* 10, 4 (2022), 78. doi:10.3390/galaxies10040078

[29] Jeffrey D. et al. Scargle. 2013. Studies in astronomical time series analysis. VI. Bayesian block representations. *The Astrophysical Journal* 764, 2 (2013), 167.

[30] Chin-Chia Michael Yeh, Yan Zhu, Liudmila Ulanova, Nurjahan Begum, Yijun Ding, Hoang Anh Dau, Diego Furtado Silva, Abdullah Mueen, and Eamonn Keogh. 2016. Matrix Profile I: All Pairs Similarity Joins for Time Series: A Unifying View that Includes Motifs, Discords and Shapelets. *IEEE ICDM* (2016). doi:10.1109/ICDM.2016.0179

[31] Bing Zhang. 2018. *The Physics of Gamma-Ray Bursts*. Cambridge University Press.

[32] S. Y. Zhu, W. P. Sun, D. L. Ma, and F. W. Zhang. 2024. Classification of Fermi gamma-ray bursts based on machine learning. *Monthly Notices of the Royal Astronomical Society* 532, 2 (2024), 1434–1443. doi:10.1093/mnras/stae1594

[33] Mansur Ziiatdinov, Farida Farsian, Francesco Schilliró, and Salvatore Distefano. 2025. Comparing Quantum Machine Learning Approaches in Astrophysical Signal Detection. In *2025 IEEE International Conference on Quantum Software*. arXiv:2507.19505 [astro-ph.IM] doi:10.1109/QSW67625.2025.00020