Published: 13 January 2026

Citation in BibTeX format

SIGSPATIAL '25: The 33rd ACM International Conference on Advances in Geographic Information Systems
November 3 - 6, 2025
MN, Minneapolis, USA

Conference Sponsors:
SIGSPATIAL

INVITED-TALK

A Randomized Data Structure for Efficient Quantum Encoding

JOHANN MAXIMILIAN ZOLLNER, Technical University of Munich, Munich, Bayern, Germany

CARLA SOPHIE RIEGER, Technical University of Munich, Munich, Bayern, Germany

MARTIN WERNER, Technical University of Munich, Munich, Bayern, Germany

Q-Spatial '25: Proceedings of the 1st ACM SIGSPATIAL International Workshop on Quantum Computing for Spatial Data Systems and Applications (November 2025)
https://doi.org/10.1145/3764923.3777404
ISBN: 9798400721861

A Randomized Data Structure for Efficient Quantum Encoding

Johann Maximilian Zollner
Big Geospatial Data Management,
School of Engineering and Design,
Technical University of Munich
maximilian.zollner@tum.de

Carla Rieger
Big Geospatial Data Management,
School of Engineering and Design,
Technical University of Munich
carla.rieger@tum.de

Martin Werner
Big Geospatial Data Management,
School of Engineering and Design,
Technical University of Munich
martin.werner@tum.de

Abstract

Near-term quantum computers are constrained by limited qubit counts and shallow circuit depth, making them unsuitable for large-scale data processing. To address this challenge, we introduce a framework for randomized data representation that enables the efficient encoding of low-cardinality classical data for quantum circuits, as presented in [12]. Using Bloom filters and fragment-based ensemble learning, our method distributes information on multiple compact quantum circuits to ensure low hardware requirements. The data structure provides the flexibility to trade representation accuracy for representation size. Our evaluation suggests that combining lossy data compression with ensemble-based quantum learning offers a promising direction toward practical quantum machine learning on near-term devices.

CCS Concepts

• **Computer systems organization → Quantum computing**; • **Information systems → Data compression**.

Keywords

Quantum Embeddings, Data Preprocessing, Quantum Classifiers, Ensemble Methods

ACM Reference Format:
Johann Maximilian Zollner, Carla Rieger, and Martin Werner. 2025. A Randomized Data Structure for Efficient Quantum Encoding. In *The 1st ACM SIGSPATIAL International Workshop on Quantum Computing for Spatial Data Systems and Applications (Q-Spatial '25), November 3–6, 2025, Minneapolis, MN, USA.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3764923.3777404

1 Introduction

Current noisy intermediate-scale quantum computers (NISQ) [7] face significant limitations in addressing large-scale data problems. These constraints are due to the limited qubit counts and low circuit depth, as well as hardware noise. As a result, quantum computing is expected to better suit problems of high complexity but small input size [4]. Hybrid quantum-classical algorithms have emerged as a promising way to leverage existing quantum hardware [1]. In such systems, classical programs offload parts of their computation to variational or parameterized quantum circuits, which can learn optimal parameters for specific tasks, such as classification.

However, efficiently encoding classical data into quantum states remains a central challenge, as state preparation circuits can be complex and often restrict the amount of information that can be encoded in practice [6]. While diverse state preparation circuits exist [10], the depth of the circuit severely limits the use of various encoding methods in practice, as increasing the circuit depth is accompanied by stronger noise effects. A low-depth quantum state preparation is, for example, basis encoding, which only relies on state preparation circuits with depth $O(1)$. Generally speaking, data is transformed on classical hardware to fit the limited input domain of a parameterized quantum circuit, and the measurement $\langle M \rangle$ of one or multiple qubits is mapped to the model's prediction y in the form of:

$$f_\theta(\phi(x)) = \langle M \rangle_{\phi(x),\theta} \mapsto y \tag{1}$$

where $\phi(x)$ describes the transformed data.

We propose a compact and lossy data representation framework that enables low-cardinality data to be effectively encoded in small-sized quantum circuits. We quantize data into representations based on a few bits and distribute fragments of this information across multiple small quantum classifiers, which are then combined into a more powerful ensemble model.

2 Method Description

Our proposed framework [12] encodes classical binarized data $x \in \{0,1\}^n$ as $\phi(x) \in \{0,1\}^m$ with $m, n \in \mathbb{Z}$ and $m < n$ using a Bloom filter, transforming raster data into a pseudo-random bit array. A Bloom filter is a compact representation for a set membership test with false positives (FPs) but no false negatives [2]. Each element's presence is probabilistically marked, yielding a representation of an input with size n with a tunable FP rate:

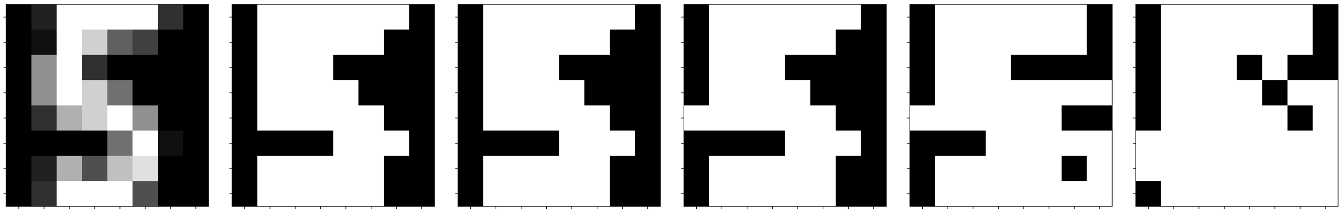
$$p = \left(1 - \exp\frac{-kn}{m}\right)^k, \tag{2}$$

for a filter with m bits and k hash functions. Hence, the data structure allows for a trade-off between the probability p and the number of bits m. Specifically, we utilize GloBiMaps, which provides possibilities for augmentation, to efficiently store a binary raster as a global bitmap [11]. Consequently, we map binary raster data x to a pseudo-random bit string of a power of two sizes m with hash functions h_i with $i = 1, ..., k$. Then, the quantum encoding of the bits b_j with single-qubit Pauli-X gates and depth $O(1)$ can be simply described as $b_j \mapsto |b_j\rangle$, denoting by $|b_j\rangle$ the computational basis states.

The impact of the parameter p is best illustrated when applying the randomized transformation to an exemplary image, as shown in Figure 1. By saving the indices (f, l) of binarized pixels $I_{f,l}$ and

Q-Spatial '25, Minneapolis, MN, USA

ACM ISBN 979-8-4007-2186-1/2025/11
https://doi.org/10.1145/3764923.3777404



(a) Original image **(b) Binarized image** **(c)** $p = 0.02$ **(d)** $p = 0.14$ **(e)** $p = 0.40$ **(f)** $p = 0.64$

Figure 1: Impact of the false positive rate p of the randomized transformation on an image [12]. Figures c) - f) show archetypes of transformed images generated using pseudo-inverted hash functions $h_i^{-1}(b_j)$ with varying configurations for b_j.

the indices of the corresponding filter elements b_j while mapping $I_{f,l} \mapsto b_j$, h_i may be pseudo-inverted to retain the binary input value. Such an inversion can be generally described as $h_i^{-1}(b_j)$ and results in an archetype for the input. It shows that it is possible to reconstruct the image error-free from a filter, while the archetypes are considerably disturbed with increasing FP rate.

In order to adapt to the limited qubit availability, the bit array is divided into fragments of size m/N, where N is the number of qubits in the quantum circuit, each representing a subset of the encoded information suitable for a small quantum circuit. Each fragment serves as input to one of $2m/N - 1$ separate quantum classifiers, and the ensemble of these collectively reconstructs the original decision boundaries. To be precise, we utilize bootstrap aggregation, or bagging, an ensemble learning method where the model takes the majority over multiple base models. Bootstrap aggregation provides improved generalization and enables resource savings in quantum machine learning [5].

3 Evaluation

Our approach is evaluated for binary classification tasks based on the handwritten digits [3] and the Iris dataset. Data were binarized, rasterized, Bloom-filter encoded, and fragmented into small bit sequences. Those were encoded with basis encoding in $N \in \{4, 8, 16\}$ quantum circuits, i.e., evaluating on the circuit-centric classifier [8] and multi-scale entanglement renormalization ansatz [9] variational circuits, while SVMs and decision trees served as classical baselines. Results show that where the theoretically calculated FP rate of the randomized representations exceeds 0.5, a significant decrease in performance can be observed. However, even weak quantum classifiers with 4 out of 512 random bits and an FP rate of up to $p = 0.64$ as input can achieve > 60 % accuracy. The overall performance of these small-sized classifiers significantly enhances with ensemble aggregation, achieving accuracies > 90 %. While the evaluated individual classifiers are outperformed by classical methods, their ensemble model demonstrates competitive performance.

Increasing the Bloom filter's FP rate only led to a minor performance decrease. Meanwhile, increasing fragment diversity, rather than qubit count, was the strongest impact on performance. This indicates that randomized encoding provides an inherent form of regularization and diversity for ensemble learning. Furthermore, the quantum circuit simulations also benefit from the data transformation, as the computing cost of simulating qubits grows exponentially with their number, whereas the computing cost increases linearly with additional base classifiers of the same qubit register size.

4 Conclusion

We presented a randomized transformation for efficient quantum data encoding using Bloom filters and fragmentation. This method enables low-cardinality data to be represented as few-bit arrays suitable for shallow quantum circuits. These representations not only simplify the design of embedded circuits but also introduce useful randomness that enhances ensemble quantum learning models. The results show that small, low-fidelity quantum representations can achieve competitive performance when combined in ensemble architectures. Overall, the study highlights the potential of classical preprocessing to bridge the gap between current hardware limitations and practical quantum machine learning applications.

References

[1] Marcello Benedetti, Erika Lloyd, Stefan Sack, and Mattia Fiorentini. 2019. Parameterized quantum circuits as machine learning models. *Quantum Science and Technology* 4, 4 (2019), 043001. doi:10.1088/2058-9565/ab4eb5

[2] Burton H Bloom. 1970. Space/time trade-offs in hash coding with allowable errors. *Commun. ACM* 13, 7 (1970), 422–426. doi:10.1145/362686.362692

[3] Li Deng. 2012. The MNIST Database of Handwritten Digit Images for Machine Learning Research [Best of the Web]. *IEEE Signal Processing Magazine* 29, 6 (2012), 141–142. doi:10.1109/MSP.2012.2211477

[4] Torsten Hoefler, Thomas Häner, and Matthias Troyer. 2023. Disentangling hype from practicality: on realistically achieving quantum advantage. *Commun. ACM* 66, 5 (2023), 82–87. doi:10.1145/3571725

[5] Massimiliano Incudini, Michele Grossi, Andrea Ceschini, Antonio Mandarino, Massimo Panella, Sofia Vallecorsa, and David Windridge. 2023. Resource saving via ensemble techniques for quantum neural networks. *Quantum Machine Intelligence* 5, 2 (2023), 39.

[6] Seth Lloyd, Maria Schuld, Aroosa Ijaz, Josh Izaac, and Nathan Killoran. 2020. Quantum embeddings for machine learning. arXiv:2001.03622 [quant-ph] https://arxiv.org/abs/2001.03622

[7] John Preskill. 2018. Quantum Computing in the NISQ era and beyond. *Quantum* 2 (Aug. 2018), 79. doi:10.22331/q-2018-08-06-79

[8] Maria Schuld, Alex Bocharov, Krysta M Svore, and Nathan Wiebe. 2020. Circuit-centric quantum classifiers. *Physical Review A* 101, 3 (2020), 032308. doi:10.1103/PhysRevA.101.032308

[9] G. Vidal. 2008. Class of Quantum Many-Body States That Can Be Efficiently Simulated. *Physical Review Letters* 101, 11 (sep 2008). doi:10.1103/PhysRevLett.101.110501

[10] Manuela Weigold, Johanna Barzen, Frank Leymann, and Marie Salm. 2021. Encoding patterns for quantum algorithms. *IET Quantum Communication* 2, 4 (2021), 141–152. doi:10.1049/qtc2.12032

[11] Martin Werner. 2019. Globimaps-A probabilistic data structure for in-memory processing of global raster datasets. In *Proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems.* 3–12. doi:10.1145/3453184

[12] Johann Maximilian Zollner and Martin Werner. 2025. A randomized transformation for data processing on quantum computers. *Quantum Machine Intelligence* 7, 1 (2025), 22.