Quantum Information Processing (2025) 24:106
https://doi.org/10.1007/s11128-025-04725-6

A Laplacian-based quantum graph neural networks for quantum semi-supervised learning

Hamed Gholipour[1,2] · Farid Bozorgnia[3] · Kailash Hambarde[1] · Hamzeh Mohammadigheymasi[1,4,5] · Javier Mancilla[6] · Andre Sequeira[7] · João Neves[1] · Hugo Proença[8] · Moharram Challenger[9,10]

Received: 19 September 2024 / Accepted: 10 March 2025

Abstract
The Laplacian learning method has proven effective in classical graph-based semi-supervised learning, yet its quantum counterpart remains underexplored. This study systematically evaluates the Laplacian-based quantum semi-supervised learning (QSSL) approach across four benchmark datasets—Iris, Wine, Breast Cancer Wisconsin, and Heart Disease. By experimenting with varying qubit counts and entangling layers, we demonstrate that increased quantum resources do not necessarily lead to improved performance. Our findings reveal that the effectiveness of the method is highly sensitive to dataset characteristics, as well as the number of entangling layers. Optimal configurations, generally featuring moderate entanglement, strike a balance between model complexity and generalization. These results emphasize the importance of dataset-specific hyperparameter tuning in quantum semi-supervised learning frameworks.

Keywords Quantum semi-supervised learning (QSLL) · Quantum graph learning · Parametrized quantum circuits · Laplacian QSSL · Entanglement · Test accuracy

1 Introduction

The development of machine learning techniques has undergone significant transformations over the past few decades, evolving from simple linear models to sophisticated deep learning architectures [1, 2]. Initially, the focus of machine learning was on supervised learning, where models were trained on fully labeled datasets to predict outcomes on unseen data [3, 4]. Supervised learning (SL) has been effectively applied in fields such as biomedical engineering, finance, and Earth and environmental sciences, providing robust solutions for predictive modeling and data analysis [5]. However, the increasing need to process large volumes of unlabeled data led to the development of unsupervised learning techniques, which are designed to discover patterns and structures in data without labeled outcomes. This evolution paved the way for Semi-Supervised Learning (SSL), a hybrid approach that combines the strengths

Extended author information available on the last page of the article

Published online: 03 April 2025

of supervised and unsupervised learning to utilize both labeled and unlabeled data, thereby improving model accuracy and robustness [6]. SSL addresses the challenge of limited labeled data by leveraging a mix of labeled and unlabeled data during training, effectively bridging the gap between supervised and unsupervised learning to enhance model performance [7].

Integrating quantum computing with machine learning represents the latest frontier in this evolutionary journey [8, 9]. One of the most innovative advancements in quantum machine learning (QML) is quantum graph learning (QGL), which synergizes quantum computing with graph-based learning methodologies [10, 11]. By incorporating quantum circuits into graph neural networks (GNNs), QGL offers a powerful framework to address complex challenges in graph learning. This novel approach holds promise for various applications, from optimizing network communications to advancing drug discovery and analysis [12]. Graph-based SSL techniques represent data points as nodes in a graph, with edges denoting relationships or similarities. These methods propagate label information across the graph to infer labels for unlabeled nodes, often using techniques like label propagation and graph-based regularization to maintain label consistency among neighboring nodes. A prominent graph-based SSL method is Laplacian learning, initially detailed by Zhu et al. [13]. While Laplacian learning has demonstrated strong performance in certain scenarios, it has limitations, particularly in classification tasks with a small number of labels [14].

This study explores the integration of quantum graph neural networks within the framework of Laplacian quantum semi-supervised learning (QSSL). It evaluates the effectiveness of Laplacian QSSL methods across four benchmark datasets: Iris [15], Wine [16], Breast Cancer Wisconsin Diagnostic [17], and Heart Disease [18]. The evaluation focuses on systematically investigating how varying the number of qubits and the number of entangling layers affect test accuracy. Notably, increasing the number of qubits does not always improve performance; this reflects a complex interplay of factors. As qubit counts rise, the quantum circuits become more complex, leading to higher error rates and potential performance degradation. A key challenge is the barren plateau phenomenon, where the cost function's gradient becomes vanishingly small as the number of qubits increases. This makes optimization difficult, as the gradient of the cost function decreases exponentially with the number of qubits. Consequently, after around 14 qubits, a noticeable degradation in the cost function is observed, indicating significant difficulties in scaling up quantum circuits. This underscores the complexity of achieving scalable quantum computing.

Moreover, quantum algorithms like the Laplacian method may not be optimized for larger qubits configurations, resulting in a disproportionate increase in computational burden. The effectiveness of additional qubits also depends on the characteristics of the dataset, as they may not always align well with the data's features. Addressing these challenges requires optimizing quantum algorithms, managing resources efficiently, and carefully tuning models to balance complexity and performance. Understanding these dynamics is crucial for advancing quantum machine learning and the benefits of quantum computing. Additionally, the performance of the Laplacian learning method is highly sensitive to the number of entangling layers, with optimal configurations varying across datasets. Generally, moderate levels of entanglement strike the

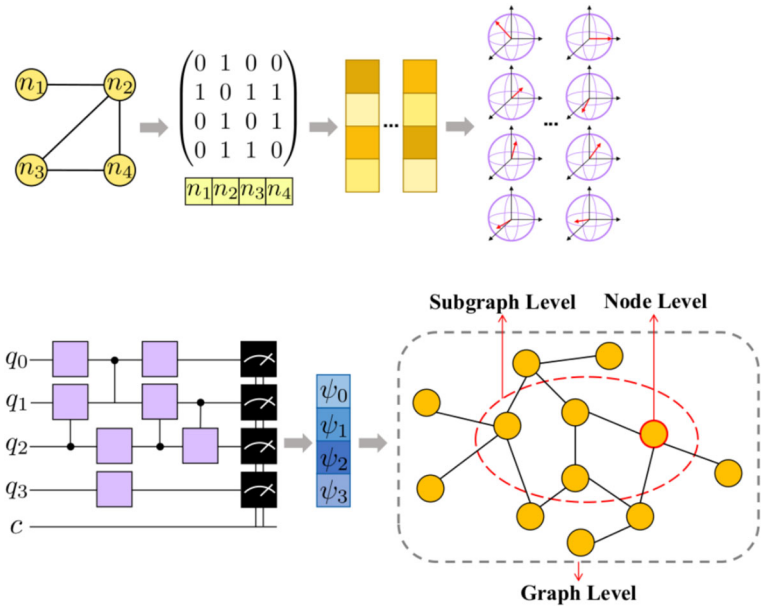

Fig. 1 A quantum circuit architecture for GNNs, including input encoding, variational quantum layers, and measurement. This hybrid quantum-classical approach aims to enhance GNNs performance in tasks such as node classification and link prediction (Figure modified from [19])

best balance between model complexity and generalization capability. These findings highlight the importance of hyperparameter tuning specific to each dataset to achieve optimal performance in Laplacian learning methods. Fully quantum workflows for graph-based machine learning hold great promise but are currently limited by the scalability of quantum algorithms for tasks such as Laplacian computation and label propagation, as well as the capabilities of existing quantum hardware. Therefore, this study takes a hybrid approach, utilizing classical graph techniques for efficient and reliable preprocessing while incorporating quantum-specific enhancements tailored to the classification stage.

This paper is structured as follows: We commence with a comprehensive overview of quantum graph learning in Sect. 2, highlighting its relevance to SSL. Subsequently, we introduce the Laplacian-based quantum semi-supervised learning approach and its quantum implementation, supported by a detailed mathematical framework and an explicit algorithm in Sect. 3. Following this, we conduct a thorough examination and performance analysis of the QSSL methodology, employing various numerical experiments across diverse datasets in Sect. 4. We then discuss the implications of modifications to qubit count and the number of entangling layers, analyzing their effects on test accuracy and quantum entanglement within the context of the Laplacian-based QSSL framework. Finally, we conclude the paper in Sect. 6 by synthesizing the key findings and contributions of our research (Fig. 1).

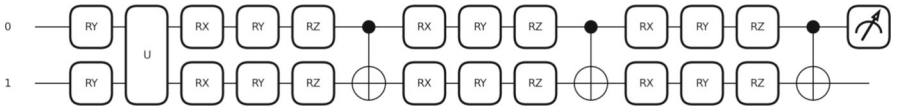

Fig. 2 The quantum circuit used in the paper, detailing its structure and components including entangling layers, unitary operations, and quantum gates

2 Background: quantum graph learning

In this section, we explore graph-based learning and its application in quantum semi-supervised learning through QGL. Graph learning focuses on algorithms that analyze data structured as graphs, with nodes representing entities and edges depicting their relationships. Graphs model complex systems such as social networks, biological interactions, and knowledge bases. A fundamental aspect is understanding graph structures, where nodes and edges can be directed, undirected, or weighted. GNNs extend traditional neural networks to handle graph-structured data by incorporating connectivity through message passing, enabling nodes to aggregate information from neighbors. Core tasks in graph learning include node classification, link prediction, graph classification, and clustering [20]. Techniques employed involve spectral methods (e.g., Laplacian matrix), spatial methods (e.g., graph convolutional networks), random walk-based methods (e.g., DeepWalk, node2vec), and attention mechanisms (e.g., Graph Attention Networks) [21]. Graph learning is applied across various domains, including social networks for interaction analysis and community detection, biological networks for understanding protein interactions, and knowledge graphs for enhancing search engines and recommendation systems [22].

QGL leverages quantum computing to address graph learning tasks and is classified into three types: quantum computing on graphs, quantum graph representation, and quantum circuits for GNNs. This paper focuses on quantum circuits for GNNs, particularly with GNNs, particularly with noisy intermediate-scale quantum (NISQ) devices, which integrate GNNs with quantum modules to enhance model performance [23] (Fig. 2).

An advantage of implementing quantum graph neural networks (QGNNs) relies on leveragingNISQ devices to adapt the structure of graph neural parametrized quantum circuits (PQCs) [24]. NISQ uses PQCs, a set of gates with free parameters to be tuned to solve a given task of interest, such as a variational quantum algorithm [25]. Quantum graph neural network is simulated by PQCs in a way that encodes input graph data in quantum amplitudes(see Figure 1 [19]). The following section will explain how quantum graph learning uses the Laplacian method in QSSL for classification.

3 Laplacian-based quantum semi-supervised learning

SSLtechniques serve as a bridge between supervised and unsupervised learning. They grasp extensive unlabeled datasets alongside a smaller set of labeled data to enhance the learning model. SSL aims to extend the known labels to unlabeled samples through a suitable smoothing operator.

3.1 Laplacian learning

Consider a graph $G = (V, E)$ with n nodes. The adjacency matrix A is an $n \times n$ matrix where A_{ij} represents the weight of the edge between nodes i and j.

The degree matrix D is a diagonal matrix where $D_{ii} = \sum_j A_{ij}$, i.e., the degree of node i.

The normalized Laplacian matrix L_{norm} is defined as:

$$L_{\text{norm}} = I - D^{-1/2} A D^{-1/2}$$

where I is the identity matrix.

Let $\mathbf{Y}$ be the initial label matrix of size $n \times c$, where n is the number of nodes and c is the number of classes.

To refine the label matrix $\mathbf{F}$, we use the iterative update rule:

$$\mathbf{F}^{(t+1)} = \alpha L_{\text{norm}} \mathbf{F}^{(t)} + (1 - \alpha)\mathbf{Y}$$

where α is a weighting factor (typically close to 1) and T is the number of iterations.

3.1.1 Iteration process

1. Initialize the label matrix:

$$\mathbf{F}^{(0)} = \mathbf{Y}$$

2. For each iteration t from 0 to $T - 1$:

$$\mathbf{F}^{(t+1)} = \alpha L_{\text{norm}} \mathbf{F}^{(t)} + (1 - \alpha)\mathbf{Y}$$

After T iterations, the refined label matrix is:

$$\mathbf{F}^{(T)}$$

3.1.2 Quantum implementation

An adjacency matrix is initially generated from a graph in the quantum realm. The QR decomposition of the adjacency matrix introduces a unique way to encode graph structure directly into the parameterized quantum circuit. This approach ensures that the quantum model can exploit graph-specific information, bridging classical graph theory and quantum processing in a novel manner. By leveraging this integration, we aim to enhance the classification accuracy compared to standard variational methods. Both Poisson and Laplacian matrices can be encoded into the quantum state. This process is referred to as amplitude encoding, where a quantum state $|\psi\rangle$ can represent the graph data:

$$|\psi\rangle = \sum_i \alpha_i |i\rangle$$

with α_i being the amplitude corresponding to node i.

Solving the Laplacian for a quantum implementation involves quantum algorithms such as the HQuantum Phase Estimation (QPE) and the Harrow–Hassidim–Lloyd (HHL) algorithm. These algorithms provide a way to efficiently solve linear systems and eigenvalue problems, which are at the core of Laplacian-based methods. For instance, the HHL algorithm can solve equations of the form $Ax = b$ efficiently, where A can be the Laplacian matrix L. The algorithm prepares a quantum state corresponding to b and then uses quantum operations to find the state corresponding to x. In practice, the adjacency matrix A, degree matrix D, and consequently the Laplacian matrix L are encoded into a quantum state. The quantum algorithm then operates on this state to find the desired solution f, which represents the labels' smooth propagation over the graph. By incorporating these mathematical details and quantum implementation aspects, the paper clarifies how the Laplacian method functions in graph-based semi-supervised learning. This addition ensures that readers can grasp these methods' theoretical underpinnings, mathematical formulation, and potential quantum advantages. As one of the semi-supervised learning methods, graph-based methods leverage the data's inherent structure, often represented as a graph, to propagate label information from labeled to unlabeled data points. These methods construct a graph where nodes represent data points, and edges capture relationships between them, such as similarity or proximity. By iteratively propagating labels through the graph, often guided by the Graph's structure or properties, graph-based semi-supervised learning algorithms can effectively generalize known labels to unlabeled samples [13]. Indeed, a significant portion of graph-based learning methods employs the graph Laplacian as the smoothing operator to facilitate generalization. However, more sophisticated nonlinear approaches resort to p-Laplacian operators for enhanced performance. These operators enable a more nuanced representation of the data, better capturing complex relationships and structures within the graph [26]. You can find more information about graph learning in two references [27, 28].

In the transition to the quantum realm, an adjacency matrix is initially generated from a graph. Subsequently, Laplacian matrices can be encoded into the quantum state or amplitude of the quantum state vector, a process referred to as amplitude encoding [29, 30].

3.2 An algorithm for Laplacian-based quantum semi-supervised learning

This algorithm integrates quantum computing with Laplacian learning to perform quantum semi-supervised classification. It utilizes quantum circuits for classification and refines predictions using Laplacian smoothing. Additionally, the algorithm includes parameter tuning and performance evaluation through visualization.

1: **Input:** Labeled data (X_l, y_l), unlabeled data X_u, parameters α, T, K, η, L, n, hyperparameter range H_{range}.
2: **Output:** Predictions $\hat{y}_u$, entanglement entropies S_u, performance metrics (Accuracy, Precision, Recall, F1 Score).
3: **Step 1: Data Preprocessing**
- Replace missing values in X:
$$X_{ij} \leftarrow \text{median}(X_{\cdot j}) \quad \text{if } X_{ij} \text{ is missing.}$$
- Binarize target labels:
$$y_i = \begin{cases} 1 & \text{if } y_i > 0, \\ 0 & \text{otherwise.} \end{cases}$$
- Standardize features:
$$X_{\text{scaled}} = \frac{X-\mu}{\sigma}.$$
- Adjust feature dimensionality to match n qubits.

4: **Step 2: Graph Construction**
- Construct adjacency matrix A using Gaussian kernel:
$$A_{ij} = \exp\left(-\frac{\|X_i - X_j\|^2}{2\sigma^2}\right).$$
- Compute degree matrix D:
$$D_{ii} = \sum_j A_{ij}.$$
- Compute normalized Laplacian matrix L_{norm}:
$$L_{\text{norm}} = I - D^{-\frac{1}{2}} A D^{-\frac{1}{2}}.$$

5: **Step 3: Laplacian Learning**
- Initialize label matrix Y:
$$Y = \begin{bmatrix} y_l \\ \mathbf{0}_{|X_u| \times 1} \end{bmatrix}.$$
- Iteratively update labels using Laplacian smoothing:
$$F^{(t+1)} = \alpha L_{\text{norm}} F^{(t)} + (1-\alpha) Y, \quad t = 1, \ldots, T.$$
- Extract refined labels for unlabeled data:
$$y_u = F^{(T)}[|X_l| :, 0].$$

6: **Step 4: Quantum Circuit Setup**
- Construct unitary matrix U via QR decomposition:
$$U = \text{QR}(A).$$
- Define quantum circuit:
  - Angle embedding of features:
  $$\text{AngleEmbedding}(X_i, \text{wires}).$$
  - Apply unitary matrix:
  $$\text{QubitUnitary}(U, \text{wires}).$$
  - Parameterized quantum layers:
  $$\text{StronglyEntanglingLayers}(\theta, L, \text{wires}).$$

7: **Step 5: Optimization**
- Define binary cross-entropy loss:
$$\mathcal{L}(\theta) = -\frac{1}{N} \sum_{i=1}^{N} \left[y_i \log(p_i) + (1-y_i) \log(1-p_i) \right],$$
where $p_i = \sigma(\text{quantum\_circuit}(X_i, \theta))$.
- Optimize parameters using gradient descent:
$$\theta^{(k+1)} = \theta^{(k)} - \eta \nabla_\theta \mathcal{L}(\theta), \quad k = 1, \ldots, K.$$

8: **Step 6: Prediction and Entropy Calculation**
- Predict labels for unlabeled data:
$$\hat{y}_i = \begin{cases} 1 & \text{if quantum\_circuit}(X_i, \theta) > 0.5, \\ 0 & \text{otherwise.} \end{cases}$$
- Compute entanglement entropy for each quantum state:

$S = -\sum_i \lambda_i \log_2(\lambda_i)$,
where λ_i are eigenvalues of the reduced density matrix $\rho = \mathrm{Tr}_{\mathrm{env}}(|\psi\rangle\langle\psi|)$.
9: **Step 7: Performance Evaluation**
• Compute metrics: Accuracy, Precision, Recall, F1 Score.
10: **Step 8: Visualization**
• Plot performance metrics and entanglement entropy distributions.
11: **Return:** Predictions $\hat{y}_u$, entropies S_u, performance metrics.

Table 1 Classification parameters, test accuracy, F1, recall and precision in four datasets Iris, Wine, Breast cancer and Heart Disease

Dataset	Laplacian			
	Test accuracy	F1	Recall	Precision
Iris	0.822	0.779	0.938	0.666
Wine	0.533	0.515	0.753	0.391
Breast cancer	0.764	0.827	0.904	0.763
Heart disease	0.487	0.474	0.500	0.450

4 Numerical experiments

4.1 Analyses of classification parameters for four datasets

The Laplacian classification method performs well on the Iris and Breast Cancer datasets, achieving high accuracy, recall, and a balanced F1 Score. Specifically, it reaches an accuracy of 82.2 and a Recall of 93.8 on the Iris dataset, indicating effective classification with a trade-off in Precision. On the Breast Cancer dataset, it achieves a 76.4 accuracy and a high F1 Score of 0.827, reflecting strong performance with both high Recall and Precision.

In contrast, the method shows weaker results for the Wine and Heart Disease datasets. The Wine dataset has an accuracy of 53.3 and a low precision of 39.1, resulting in a lower F1 Score of 0.515. The Heart Disease dataset presents the greatest challenge, with an accuracy of only 48.7, a low precision of 45.0, and a modest recall of 50.0, leading to the lowest F1 score of 0.474.

Overall, the Laplacian method is effective for simpler datasets like Iris and Breast Cancer. However, it struggles with more complex datasets such as Wine and Heart Disease, showing lower accuracy and higher rates of false positives (Table 1).

4.2 Analyses of classification parameters by changing qubits

The performance of the Laplacian method across different datasets and qubit configurations reveals several key insights. For the Iris dataset, the model performs optimally with four qubits, achieving a test accuracy of 0.822, an F1 score of 0.779, a recall of 0.938, and a precision of 0.666. In contrast, increasing the number of qubits to 8, 12, and 14 significantly degrades the model's performance, with accuracy dropping to 0.352 and precision falling to 0.263 for the 14-qubit configuration, indicating diminishing returns or inefficiency with additional quantum resources.

Similarly, for the Wine dataset, the Laplacian method yields its highest test accuracy of 0.533 and an F1 score of 0.515 with four qubits. As qubits increase, performance deteriorates, with the 14-qubit configuration showing the lowest accuracy at 0.460 and an F1 Score of 0.313. This suggests that the method struggles to leverage additional qubits effectively for this dataset.

For the Breast Cancer dataset, the Laplacian method also performs best with four qubits, achieving a test accuracy of 0.764 and an F1 score of 0.827. While performance remains relatively high with eight qubits, accuracy and F1 Score drop significantly with 12 and 14 qubits, reaching 0.447 and 0.264, respectively. This decline suggests that beyond a certain point, the added complexity of more qubits hampers the model's generalization capability.

In the Heart Disease dataset, the Laplacian method achieves the best results with 4 qubits, with a test accuracy of 0.487 and an F1 Score of 0.474. Increasing the number of qubits leads to diminishing returns, with minimal improvement at 12 and 14 qubits, underscoring inefficiency in using additional quantum resources for this dataset.

These results demonstrate that increasing the number of qubits does not consistently improve the Laplacian method's performance. The optimal qubit count varies by dataset, and additional qubits often introduce complexity without yielding better performance. This observation underscores the need for careful tuning of quantum resources, suggesting that fewer qubits may be more effective in certain cases. As qubit counts increase, quantum circuits become more intricate, potentially leading to higher error rates and degraded performance. Furthermore, certain quantum algorithms like the Laplacian method may not be well-suited to larger qubit configurations, increasing computational burdens disproportionately. The characteristics of each dataset also influence performance, as additional qubits may not align well with the data's features. Thus, optimizing quantum algorithms, managing resources efficiently, and balancing complexity with performance are crucial for advancing quantum machine learning (Fig. 3, Tables 2, 3, 4).

4.3 Analysis of entanglement and test accuracy by changing entangling layers

Studying entanglement as a fundamental quantum feature of our quantum systems is precious and promising in this paper. By investigating the impact of entangling layers on the entanglement properties of the systems and classification parameters, such as test accuracy, we can gain deeper insights into the mechanisms that drive the performance of quantum algorithms. This understanding can inform the design of more effective quantum circuits and enhance our ability to utilize quantum entanglement for improved computational tasks. The table presents an analysis of entanglement based on the number of entangling layers in the quantum circuit, using entanglement entropy as the metric. Entanglement is generated by combining different quantum logics, such as CNOT with Z gates. The derivation of entanglement entropy starts with a quantum system described by a density matrix ρ, encompassing multiple subsystems A and B. To quantify the entanglement between subsystem A and the rest of the system B, we compute the reduced density matrix ρ_A by tracing out the degrees of freedom of subsystem B from ρ. The entanglement entropy S_A for subsystem A is then defined

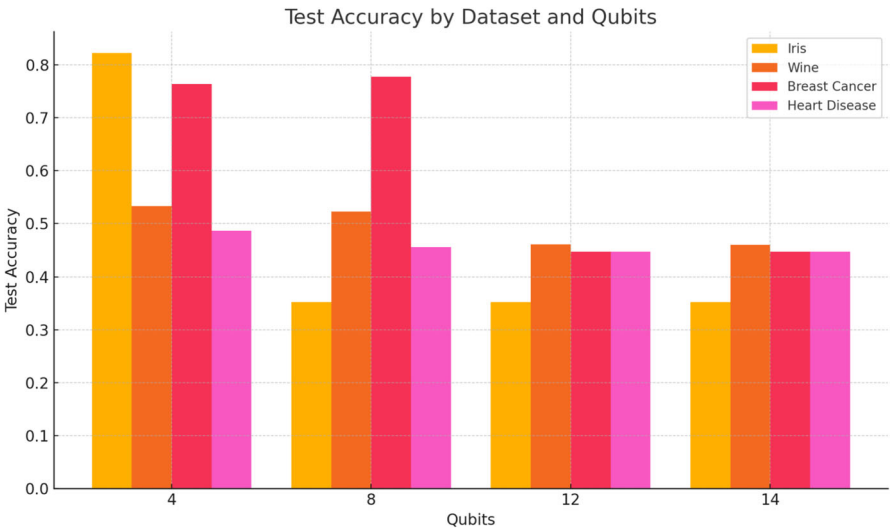

Fig. 3 Line chart showing the change in test accuracy with increasing qubits for each dataset. Each line represents a different dataset

Table 2 Quantification of classification parameters, test accuracy, F1 score, recall, and precision for the Laplacian method across four datasets (Iris, Wine, Breast Cancer, Heart Disease) with varying qubits (4, 8, 12, 14)

Dataset-qubit	Laplacian			
	Test accuracy	F1	Recall	Precision
Iris-4	0.822	0.779	0.938	0.666
Iris-8	0.352	0.350	0.525	0.350
Iris-12	0.352	0.350	0.525	0.263
Iris-14	0.352	0.350	0.525	0.263
Wine-4	0.533	0.515	0.753	0.391
Wine-8	0.523	0.486	0.687	0.376
Wine-12	0.461	0.453	0.679	0.340
Wine-14	0.460	0.313	0.383	0.264
Breast Cancer-4	0.764	0.827	0.904	0.763
Breast Cancer-8	0.778	0.837	0.909	0.776
Breast Cancer-12	0.447	0.284	0.383	0.572
Breast Cancer-14	0.447	0.264	0.383	0.568
Heart Disease-4	0.487	0.474	0.500	0.450
Heart Disease-8	0.456	0.288	0.340	0.250
Heart Disease-12	0.447	0.264	0.383	0.540
Heart Disease-14	0.447	0.264	0.383	0.539

It illustrates how the method performs as quantum resources increase, highlighting the highest metrics for each dataset and qubits configuration

Table 3 Classification metrics for semi-supervised learning methods (Iris Dataset)

Method	Accuracy	Precision	Recall	F1 Score
Self-training	0.905	0.910	0.905	0.907
Co-training	0.895	0.900	0.895	0.897
Label propagation	0.910	0.915	0.910	0.912
Label spreading	0.900	0.905	0.900	0.902
S3VM	0.880	0.885	0.880	0.882

using the von Neumann entropy formula:

$$S_A = -\mathrm{Tr}(\rho_A \log \rho_A)$$

which measures the amount of entanglement or quantum correlations between A and B. This entropy is a fundamental measure in quantum information theory, providing insights into the structure and distribution of quantum entanglement within multipartite quantum systems.

4.3.1 Measure of entanglement

In quantum computing, accurately measuring entanglement is essential for evaluating the coherence and correlations among qubits within a quantum circuit. Various metrics are employed for this purpose, each suited to different quantum systems and contexts:

Entanglement entropy This measure evaluates the entropy of the reduced density matrix of a subsystem within an entangled state. It is particularly useful for assessing the overall entanglement within a quantum system. Entanglement entropy has been selected for our analysis of the entanglement measurements discussed. This measure is preferred because of its ability to provide a comprehensive quantification of the entanglement within the entire quantum system and specific subsystems. This choice is integral to understanding the extent and distribution of entanglement among the qubits in our quantum circuit.

4.4 Analysis of the impact of entangling layers on entanglement entropy and test accuracy

The table evaluates the impact of increasing the number of entangling layers on entanglement entropy and test accuracy across four datasets (Iris, Wine, Breast Cancer, and Heart Disease). Below is a detailed analysis of the results, focusing on the interplay between entanglement entropy and test accuracy.

Iris dataset The Iris dataset exhibits a nuanced relationship between the number of entangling layers and model performance. With five layers, the model achieves a test accuracy of 0.821 and an entanglement entropy of 0.152, indicating a strong

Table 4 Evaluation of the Laplacian learning method, focusing on how increasing entangling layers affects entanglement entropy and test accuracy

Dataset	Iris						Wine						Breast cancer						Heart disease					
	5	10	15	20	25	30	5	10	15	20	25	30	5	10	15	20	25	30	5	10	15	20	25	30
Entanglement entropy	0.152	0.264	0.155	0.149	0.205	0.172	0.135	0.185	0.291	0.124	0.190	0.207	0.202	0.182	0.130	0.218	0.114	0.204	0.155	0.223	0.190	0.182	0.233	0.248
Test Accuracy	0.821	0.225	0.765	0.882	0.806	0.700	0.537	0.229	0.307	0.645	0.640	0.416	0.764	0.580	0.431	0.354	0.662	0.382	0.491	0.433	0.446	0.584	0.551	0.451

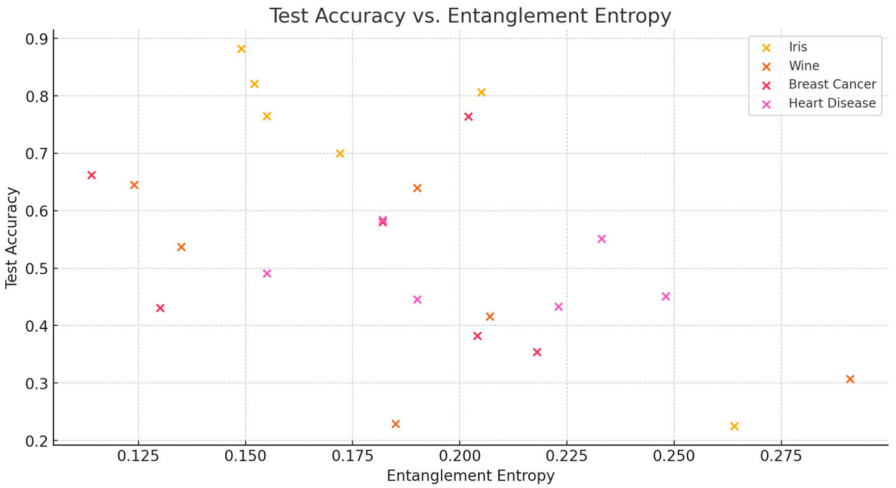

Fig. 4 The scatter plot shows the relationship between test accuracy and entanglement entropy for each dataset. There does not appear to be a strong correlation between entanglement entropy and test accuracy across the datasets

performance with low complexity. Increasing the layers to 10 results in a substantial rise in entanglement entropy to 0.264 and a sharp drop in accuracy to 0.225, suggesting overfitting or excessive complexity (Fig. 4).

However, at 15 layers, the model's performance partially recovers, with an accuracy of 0.765 and a reduced entropy of 0.155. The best performance is observed with 20 layers, where the model achieves its highest accuracy of 0.882 and the lowest entropy of 0.149. This indicates an optimal balance between model complexity and generalization. Further increasing the layers to 25 and 30 results in fluctuating entanglement entropy (0.205 and 0.172, respectively) and a decline in accuracy (0.806 and 0.700), highlighting diminishing returns and potential overfitting with excessive layers.

Wine dataset The Wine dataset shows variability in performance with changing entangling layers. The model achieves moderate accuracy (0.537) and low entropy (0.135), starting with five layers. Doubling the layers to 10 leads to a significant increase in entropy (0.185) and a drastic drop in accuracy to 0.229, indicating overfitting. At 15 layers, the entropy peaks at 0.291 while accuracy slightly improves to 0.307. Optimal performance is observed with 20 layers, achieving the highest accuracy of 0.645 and the lowest entropy of 0.124. This suggests that moderate entangling layers facilitate effective learning for the Wine dataset. Increasing the layers to 25 and 30 results in a mixed performance, with entropy values of 0.190 and 0.207 and accuracy dropping to 0.640 and 0.416, respectively. This trend highlights the adverse impact of excessive complexity.

Breast cancer dataset The optimal number of entangling layers for the Breast Cancer dataset appears minimal. With five layers, the model achieves its highest accuracy of 0.764 and an entanglement entropy of 0.202. Increasing the layers to 10 and 15

reduces accuracy (0.580 and 0.431) and lower entanglement entropy (0.182 and 0.130), indicating that additional layers do not contribute positively. At 20 layers, entropy rises to 0.218 while accuracy drops to 0.354, further confirming the inefficacy of increased layers. A slight recovery is observed with 25 layers, where entropy is minimized to 0.114, and accuracy improves to 0.662. However, increasing to 30 layers results in higher entropy (0.204) and lower accuracy (0.382), suggesting overfitting and reduced model performance with excessive layers.

Heart disease dataset The Heart Disease dataset demonstrates a complex response to the number of entangling layers. Initially, with five layers, the model achieves moderate performance with an accuracy of 0.491 and an entropy of 0.155. Increasing to 10 layers, entropy rises to 0.223, and accuracy decreases to 0.433, indicating a negative impact of additional layers. At 15 layers, entropy slightly decreases to 0.190, with a minor improvement in accuracy to 0.446. The optimal configuration is observed at 20 layers, where the model achieves the highest accuracy of 0.584 and the lowest entropy of 0.182, indicating an effective balance between complexity and performance. Further increasing the layers to 25 and 30 results in higher entropy (0.233 and 0.248) and fluctuating accuracy (0.551 and 0.451), underscoring the detrimental effect of excessive layers.

4.4.1 General observation

Relationship between entanglement entropy and test accuracy The data indicate a nonlinear and dataset-dependent relationship between entanglement entropy and test accuracy. While moderate entanglement often correlates with a higher accuracy, no consistent trend exists across all datasets. This suggests that the optimal balance between entanglement entropy and model performance must be determined empirically for each dataset.

Optimal layer configuration The optimal number of entangling layers varies significantly between datasets. For the Iris and Heart Disease datasets, 20 layers provided the best performance, yielding the highest test accuracy with relatively low entanglement entropy. In contrast, the Wine dataset performed best with 20 layers, but the Breast Cancer dataset achieved its highest accuracy with only five layers. This variability underscores the necessity of dataset-specific hyperparameter tuning to achieve optimal results.

Impact of excessive layers Increasing the number of entangling layers beyond an optimal point generally leads to overfitting. This is evidenced by a rise in entanglement entropy and a corresponding drop in test accuracy. For instance, test accuracy significantly decreased in the Iris and Wine datasets when the number of layers exceeded 20. Similarly, the Breast Cancer and Heart Disease datasets exhibited reduced accuracy with excessive layers. These findings highlight the importance of avoiding over-complexity in model design.

Moderate entanglement levels Across most datasets, moderate levels of entanglement were associated with better performance. This suggests that while some degree of entanglement is beneficial for capturing complex data patterns, excessive entanglement can hinder the model's ability to generalize, leading to decreased test accuracy.

4.4.2 Summary

The evaluation reveals that the Laplacian learning method's performance is highly sensitive to the number of entangling layers, with optimal configurations varying across datasets. Moderate entanglement levels generally balance model complexity and generalization capability best. These findings emphasize the critical importance of hyperparameter tuning tailored to each dataset to achieve optimal performance in Laplacian learning methods.

5 Discussion

5.1 Challenges of scaling quantum resources

Our results reveal that increasing quantum resources, such as qubit counts and entangling layers, does not inherently lead to improved performance. This counterintuitive finding can be attributed to several interrelated factors. First, the barren plateau phenomenon becomes more pronounced as the number of qubits increases. In this regime, the gradients of the cost function diminish exponentially, making the optimization process significantly more challenging. Although the quantum model becomes theoretically more expressive, it often struggles to converge effectively during training, particularly in larger quantum systems.

Second, noise accumulation in NISQ devices exacerbates the issue. As the complexity of quantum circuits grows with the addition of qubits and layers, error rates increase, which significantly affects computational fidelity. This is particularly problematic in current quantum hardware, where gate fidelities are not yet sufficient to fully mitigate these errors.

Third, the risk of overfitting escalates with larger quantum circuits, especially for simpler datasets. In such cases, the model tends to learn noise or irrelevant features instead of capturing the underlying data structure, resulting in a decline in generalization performance. For example, our experiments on the Iris dataset showed that increasing qubits beyond the optimal configuration resulted in a significant drop in test accuracy.

Lastly, the dataset-specific nature of quantum learning plays a crucial role. Factors such as dimensionality, class separability, and graph structure determine how effectively additional quantum resources can be utilized. In our study, moderate quantum configurations often provided the best balance between model complexity and generalization. For instance, while the Breast Cancer dataset benefited from relatively shallow circuits, the Iris dataset required a higher number of entangling layers to achieve optimal accuracy.

5.2 Generalization of findings

Although this study focuses on the Laplacian-based quantum semi-supervised learning (QSSL) framework, the challenges observed with scaling quantum resources are likely applicable to other quantum machine learning methods. Variational quantum algorithms and hybrid quantum-classical frameworks also encounter similar issues, including barren plateaus and noise sensitivity. Our findings emphasize the importance of aligning quantum model design with the specific characteristics of the dataset and the limitations of the hardware. In particular, the interplay between entanglement and circuit depth emerges as a critical factor. While moderate levels of entanglement are often beneficial for capturing complex data patterns, excessive entanglement can introduce unnecessary complexity and hinder performance. These results suggest that future research should prioritize optimizing quantum resource configurations to achieve a balance between expressiveness and generalization.

5.3 Future directions

To address the challenges identified, several avenues for future research are recommended. Advanced initialization techniques could be explored to mitigate barren plateaus and improve optimization. Additionally, error mitigation strategies tailored to NISQ devices could reduce the impact of noise in larger circuits. Finally, extending this work to a broader range of datasets with varying structural properties could provide further insights into the generalizability of these findings. Future work should also explore fully quantum workflows for graph-based learning tasks. This includes developing quantum algorithms for Laplacian computation, label propagation, and spectral graph techniques. Integrating approaches such as quantum kernel estimation or quantum graph embeddings could enable novel hybrid or fully quantum paradigms, addressing the preprocessing steps that currently rely on classical methods. These advancements would complement the findings of this study, pushing the boundaries of quantum machine learning. Our current results demonstrate the potential of Laplacian-based quantum semi-supervised learning (QSSL); however, performance trends are heavily influenced by design choices, such as the parametrized quantum circuit (PQC) ansatz and the Gaussian adjacency kernel. To further validate and generalize these results, future work will explore alternative ansatz designs, including hardware-efficient and problem-specific circuits. Additionally, the exploration of diverse kernel functions, such as polynomial and exponential kernels, will be undertaken to enhance versatility. To ensure the robustness of the proposed method, comprehensive statistical analyses will be performed across various datasets and configurations, paving the way for more effective and scalable quantum machine learning techniques.

6 Conclusion

In conclusion, increasing the number of qubits does not inherently enhance quantum computing performance, as it introduces greater complexity and higher error rates. The scalability of quantum algorithms, such as the Laplacian-based method, is often

constrained by the disproportionate computational burden associated with additional qubits. Effective alignment of qubits with dataset features remains critical, highlighting the importance of optimization, resource management, and hyperparameter tuning in quantum machine learning. The Laplacian learning method demonstrates particular sensitivity to the number of entanglement layers, with moderate levels of entanglement striking an optimal balance between computational complexity and model accuracy. A notable outcome of this study is achieving near-classical accuracy on the Iris dataset using 20 entangling layers, showcasing the potential of tailored quantum parameterization. Customizing these parameters to specific datasets is essential for optimizing performance and unlocking the full potential of quantum machine learning. While the Laplacian-based quantum semi-supervised learning (QSSL) method shows significant promise, its scalability and generalizability remain key challenges. Future research will focus on addressing these limitations by exploring alternative parametrized quantum circuit (PQC) designs, integrating diverse kernel functions, and leveraging advanced optimization techniques to mitigate trainability issues. By conducting rigorous statistical analyses and validating scalability, we aim to further enhance the robustness and applicability of this approach, paving the way for more effective quantum classification techniques capable of solving classical problems efficiently.

Acknowledgements The authors thank SociaLab and the University of Beira Interior for fostering an excellent research environment. Also, they thank Cristiano Patrício for his help formatting this manuscript's figures and tables. Additionally, we extend our gratitude to Jon Fath, the CEO of Rauva Company, for the generous financial support that contributed to the successful completion of this study. The participation of Hugo Proença in this work was funded by FCT/MCTES through national funds and by EU funds under the project UIDB/50008/2020.

Author Contributions Hamed Gholipour did model conceptualization, project administration, discussion of the results, writing and review of the manuscript; Farid Bozorgnia done model conceptualization, discussion of the results; Kailash Hambarde and Hamzeh Mohammadigheymasi developed the algorithms and reviewed the manuscript; Javier Mancilla made model conceptualization and discussion of the results; Andre Sequeira reviewed the manuscript; Joao Neves and Hugo Proenca supervised the study.

Funding Open access funding provided by FCT|FCCN (b-on).

Data availability No datasets were generated or analyzed during the current study.

Declarations

Conflict of interest The authors declare no conflict of interest.

References

1. Mahesh, B.: Machine learning algorithms-a review. Int. J. Sci. Res. **9**(1), 381–386 (2020)
2. Macaluso, A., Klusch, M., Lodi, S., Sartori, C.: Maqa: a quantum framework for supervised learning. Quantum Inf. Process. **22**(3), 159 (2023)
3. Mohammadigheymasi, H., Shi, P., Tavakolizadeh, N., Xiao, Z., Mousavi, S.M., Matias, L., Pourvahab, M., Fernandes, R.: Ipiml: A deep-scan earthquake detection and location workflow integrating pair-input deep learning model and migration location method. IEEE Trans. Geosci. Remote Sens. **61**, 1–9 (2023). https://doi.org/10.1109/TGRS.2023.3293914
4. Chapelle, O., Scholkopf, B., Zien, A (2009) Semi-supervised learning (chapelle, o. et al., eds.; 2006) [book reviews]. IEEE Trans. Neural Netw. **20**(3), 542–542 https://doi.org/10.1109/TNN.2009.2015974
5. Mohammadigheymasi, H., Tavakolizadeh, N., Matias, L., Mousavi, S.M., Silveira, G., Custódio, S., Dias, N., Fernandes, R., Moradichaloshtori, Y.: Application of deep learning for seismicity analysis in Ghana. Geosyst. Geoenviron. **2**(2), 100152 (2023)
6. Aouedi, O., Piamrat, K., Muller, G., Singh, K.: Fluids: Federated learning with semi-supervised approach for intrusion detection system. In: 2022 IEEE 19th annual consumer communications & networking conference (CCNC), pp. 523–524 (2022). https://doi.org/10.1109/CCNC49033.2022.9700632
7. Qi, H., Wang, L., Gong, C., Gani, A.: A survey on quantum data mining algorithms: challenges, advances and future directions. Quantum Inf. Process. **23**(3), 74 (2024)
8. Bhowmik, B.: Quantum machine learning and recent advancements. In: 2023 International conference on artificial intelligence and smart communication (AISC), pp. 206–211 (2023). https://doi.org/10.1109/AISC56616.2023.10085586
9. Allende, C., Olivera, A.F., Buksman, E.: Synthesis of quantum circuits based on supervised learning and correlations. Quantum Inf. Process. **23**(6), 204 (2024)
10. Medisetty, P., Kumar Pallapothu, L.K., Chand Evuru, P., Prakash, K.B., Vulavalapudi, V.M.S., Varma, G.P.S.: The quantum graph recurrent neural network. In: 2023 Second international conference on smart technologies for smart nation (SmartTechCon), pp. 667–672 (2023). https://doi.org/10.1109/SmartTechCon57526.2023.10391361
11. Hou, Y., Li, J., Chen, X., Li, H., Li, C., Tian, Y., Li, L., Cao, Z., Wang, N.: Quantum algorithm for help-training semi-supervised support vector machine. Quantum Inf. Process. **19**(9), 278 (2020)
12. Mishra, N., Kapil, M., Rakesh, H., Anand, A., Mishra, N., Warke, A., Sarkar, S., Dutta, S., Gupta, S., Prasad Dash, A., et al.: Quantum machine learning: A review and current status. Data Management, Analytics and Innovation: Proceedings of ICDMAI 2020, Volume 2, 101–145 (2021)
13. Zhu, X., Ghahramani, Z., Lafferty, J.D.: Semi-supervised learning using gaussian fields and harmonic functions. In: Proceedings of the 20th international conference on machine learning (ICML-03), pp. 912–919 (2003)
14. Flores, M., Calder, J., Lerman, G.: Analysis and algorithms for ℓ_p-based semi-supervised learning on graphs. Appl. Comput. Harmon. Anal. **60**, 77–122 (2022)
15. Fisher, R.A.: Iris. UCI Machine learning repository. https://doi.org/10.24432/C56C76 (1988)
16. Aeberhard, S., Forina, M.: Wine. UCI machine learning repository. https://doi.org/10.24432/C5PC7J (1991)
17. Wolberg, William, M., Olvi, S., Nick, S., W.: Breast Cancer Wisconsin (Diagnostic). UCI Machine Learning Repository. https://doi.org/10.24432/C5DW2B (1995)
18. Janosi, Andras, S., William, P., Matthias, D., Robert: Heart Disease. UCI Machine Learning Repository. https://doi.org/10.24432/C52P4X (1988)
19. Yu, S., Peng, C., Wang, Y., Shehzad, A., Xia, F., Hancock, E.R.: Quantum graph learning: Frontiers and outlook. arXiv preprint arXiv:2302.00892 (2023)
20. Errica, F., Podda, M., Bacciu, D., Micheli, A.: A fair comparison of graph neural networks for graph classification. arXiv preprint arXiv:1912.09893 (2019)
21. Bronstein, M.M., Bruna, J., LeCun, Y., Szlam, A., Vandergheynst, P.: Geometric deep learning: going beyond euclidean data. IEEE Signal Process. Mag. **34**(4), 18–42 (2017)
22. Skuzinski, T.: Book review of: Sotarauta, m.; beer, a.(eds.)(2021): Handbook on city and regional leadership. Raumforschung und Raumordnungl Spatial Research and Planning 80(4), 500–501 (2022)
23. Wang, M., Zheng, D., Ye, Z., Gan, Q., Li, M., Song, X., Zhou, J., Ma, C., Yu, L., Gai, Y., et al.: Deep graph library: A graph-centric, highly-performant package for graph neural networks. arXiv preprint arXiv:1909.01315 (2019)

24. Chen, Y., Wang, C., Guo, H., et al.: Novel architecture of parameterized quantum circuit for graph convolutional network. arXiv preprint arXiv:2203.03251 (2022)
25. Cerezo, M., Arrasmith, A., Babbush, R., Benjamin, S.C., Endo, S., Fujii, K., McClean, J.R., Mitarai, K., Yuan, X., Cincio, L., et al.: Variational quantum algorithms. Nat. Rev. Phys. **3**(9), 625–644 (2021)
26. Elmoataz, A., Desquesnes, X., Toutain, M.: On the game p-laplacian on weighted graphs with applications in image processing and data clustering. Eur. J. Appl. Math. **28**(6), 922–948 (2017)
27. Calder, J., Cook, B., Thorpe, M., Slepcev, D.: Poisson learning: Graph based semi-supervised learning at very low label rates. In: International conference on machine learning, pp. 1306–1316 (2020). PMLR
28. Streicher, O., Gilboa, G.: Graph laplacian for semi-supervised learning. In: International conference on scale space and variational methods in computer vision, pp. 250–262 (2023). Springer
29. Setia, K., Whitfield, J.D.: Bravyi-kitaev superfast simulation of electronic structure on a quantum computer. The Journal of chemical physics **148**(16) (2018)
30. Lloyd, S.: Quantum algorithm for solving linear systems of equations. In: APS March Meeting Abstracts, vol. 2010, pp. 4–002 (2010)

Authors and Affiliations

Hamed Gholipour[1,2] · Farid Bozorgnia[3] · Kailash Hambarde[1] · Hamzeh Mohammadigheymasi[1,4,5] · Javier Mancilla[6] · Andre Sequeira[7] · João Neves[1] · Hugo Proença[8] · Moharram Challenger[9,10]

✉ Hamed Gholipour
hamed.gholipour@ubi.pt

Farid Bozorgnia
f.bozorgnia@newuu.uz

Kailash Hambarde
kailas.srt@gmail.com

Hamzeh Mohammadigheymasi
hamzeh@g.ecc.u-tokyo.ac.jp

Javier Mancilla
javier@falcondale.pro

Andre Sequeira
andre.m.sequeira@inesctec.pt

João Neves
jcneves@ubi.pt

Hugo Proença
hugomcp@ubi.pt

Moharram Challenger
moharram.challenger@uantwerpen.be

1 Department of Computer Science, University of Beira Interior, Covilhã, Portugal

2 RAUVA Company, Lisbon, Portugal

3 Department of Mathematics, New Uzbekistan University, Tashkent, Uzbekistan

4 Atmosphere and Ocean Research Institute, The University of Tokyo, Kashiwa, Japan

5 Department of Earth and Planetary Sciences, Harvard University, Cambridge, USA

6 Falcolande Company, Vigo, Spain

7 Department of Informatics, High Assurance Software Laboratory, INESC TEC, Braga, Portugal

8 Instituto de Telecomunicações, University of Beira Interior, Covilhã, Portugal

9 Department of Computer Science, University of Antwerp, Antwerp, Belgium

10 AnSyMo/Cosys Core lab, Flanders Make Strategic Research Center, Leuven, Belgium