SSL4Q: Semi-Supervised Learning of Quantum Data with Application to Quantum State Classification

Yehui Tang [1] **Nianzu Yang** [1] **Mabiao Long** [1] **Junchi Yan** [1]

Abstract

The accurate classification of quantum states is crucial for advancing quantum computing, as it allows for the effective analysis and correct functioning of quantum devices by analyzing the statistics of the data from quantum measurements. Traditional supervised methods, which rely on extensive labeled measurement outcomes, are used to categorize unknown quantum states with different properties. However, the labeling process demands computational and memory resources that increase exponentially with the number of qubits. We propose SSL4Q, manage to achieve (for the first time) semi-supervised learning specifically designed for quantum state classification. SSL4Q's architecture is tailored to ensure permutation invariance for unordered quantum measurements and maintain robustness in the face of measurement uncertainties. Our empirical studies encompass simulations on two types of quantum systems: the Heisenberg Model and the Variational Quantum Circuit (VQC) Model, with system size reaching up to 50 qubits. The numerical results demonstrate SSL4Q's superiority over traditional supervised models in scenarios with limited labels, highlighting its potential in efficiently classifying quantum states with reduced computational and resource overhead.

1. Introduction

The classification of quantum states is a fundamental process for extracting classical information from quantum systems, essential for the certification and verification of quantum devices (Eisert et al., 2020; Kliesch & Roth, 2021). Quantum systems, when subjected to noise, often deviate from their expected evolutionary trajectories, resulting in output states with significant discrepancies. This deviation necessitates the development of a reproducible protocol for accurately classifying quantum states, which in turn facilitates benchmarking and comparison across different quantum systems. By analyzing the outcomes of quantum measurements, this protocol aims to ensure consistent classification regardless of inherent system variabilities. This approach is crucial for advancing the practical application and reliability of quantum technologies. For a comprehensive understanding, an illustrative overview is provided in Fig. 1.

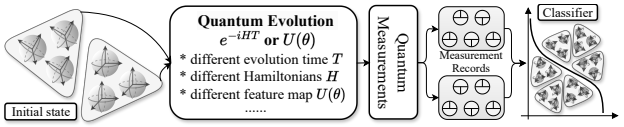

Figure 1. **An Overview of Quantum State Classification.** Given the quantum evolution specified by Hamiltonian H and time step T or by unitary feature map $\mathbf{U}(\theta)$, the quantum system initiates from state $|\psi_s\rangle$ and evolves to the final state $|\psi_f\rangle = e^{-\boldsymbol{i}HT}|\psi_s\rangle$ or $|\psi_f\rangle = \mathbf{U}(\theta)|\psi_s\rangle$ in either physical or simulated experiments. Measurements are then performed on this final state, generating records that are stored in classical memory (referred to as quantum data). Classification involves the statistical analysis of these records to categorize $|\psi_f\rangle$ into different groups. This is achieved by deducing properties of the state, such as coupling strengths of H, and the category of quantum feature maps.

The classification of quantum states hinges on the effective classical representation of these states. However, this task is fraught with challenges, primarily due to the exponential complexity inherent in quantum many-body systems (Nielsen & Chuang, 2010). Despite this complexity, it is not an insurmountable barrier. It is crucial to recognize that physical systems, especially those governed by the dynamics of local Hamiltonians, are not generic. Their intrinsic structural properties suggest that the full complexity of the Hilbert space is not always necessary for their accurate description (Carrasquilla et al., 2019).

There has been a burgeoning effort dedicated to developing efficient methodologies for quantum state classification. These methods encompass various aspects, ranging from the

[1]School of Artificial Intelligence & Department of Computer Science and Engineering & MoE Lab of AI, Shanghai Jiao Tong University, Shanghai, China. Correspondence to: Junchi Yan <yanjunchi@sjtu.edu.cn>.

Proceedings of the 41^{st} International Conference on Machine Learning, Vienna, Austria. PMLR 235, 2024.

fidelity estimation with respect to the reconstructed wave functions (Jullien et al., 2014; Flammia et al., 2012; Hibat-Allah et al., 2020; Cha et al., 2021; Hu et al., 2023) to the properties estimation of quantum states (Huang et al., 2020; 2022; Wu et al., 2023b; Lewis et al., 2023). The primary objective of these approaches is to enhance the precision of state characterization while minimizing the need for multiple copies of the quantum states. This work focuses on developing a discriminative model for categorizing quantum states with respect to specific properties such as coupling strengths, without the need to reconstruct the density representations. This can be formulated as a classification task to predict the categories of unknown quantum states.

Recently, there has been a surge in learning-based methods for addressing the challenges of learning classical representations of quantum states (Carleo et al., 2019; Carrasquilla, 2020; Karagiorgi et al., 2022; Xiao et al., 2022; Miles et al., 2023; Zhang & Di Ventra, 2023; Tang et al., 2024), which are instrumental for quantum state classification. These methods stand in contrast to traditional learning-free approaches (Burgarth & Yuasa, 2012; Gammelmark & Mølmer, 2014; Zhang & Sarovar, 2015), which typically address these problems through optimization techniques. Instead, learning-based methods typically reframe the problem as a supervised learning task (Gebhart et al., 2023). The versatility of neural networks enables the concurrent modeling of multiple systems, uncovering latent features across diverse quantum data (Zhang & Di Ventra, 2023; Wu et al., 2023b). Moreover, these methods often surpass traditional learning-free approaches in predictive accuracy and sample complexity for unknown quantum systems, a crucial advantage in resource-intensive quantum physics experiments (Zhang et al., 2021b).

A significant challenge inherent to supervised approaches is the assumption that an extensive, labeled quantum dataset is available (Gebhart et al., 2023). This premise rarely holds true in real-world scenarios. Labeling quantum data necessitates meticulous statistical analysis of measurement outcomes from (simulated) experiments (D'Ariano et al., 2003). This process not only incurs considerable experimental costs but also relies heavily on substantial classical computing resources. Moreover, the computational complexity tends to increase exponentially with the size of the quantum system (Jullien et al., 2014). A more realistic scenario, however, is one where a significant volume of quantum measurement data is available, but only a fraction of it is labeled. This situation closely mirrors the challenges faced in traditional semi-supervised learning (Van Engelen & Hoos, 2020). Despite this, the potential of semi-supervised learning in the context of quantum physics remains largely untapped.

To fill this gap, we focus on the development of a semi-supervised learning model specifically tailored for quantum data in the application of quantum state classification. By doing so, we aim to fully harness the potential of the available label information while concurrently extracting valuable insights from the expansive pool of unlabeled quantum data. Although there is a plethora of semi-supervised learning techniques available (Van Engelen & Hoos, 2020; Chen et al., 2022; Søgaard, 2022), their direct application to the classification of quantum states poses distinct challenges. In this paper, we introduce the SSL4Q, especially designed to excel in scenarios where quantum data is sparsely labeled. SSL4Q is structured to overcome four obstacles: **1) adaptability to any number of measurements; 2) permutation invariance for unordered measurements; 3) robustness against measurement uncertainties and 4) effective exploitation of unlabeled data for regularization.**

The contributions of this paper are: 1) We introduce two novel methods: shadow embedding and the PEA (Permutation Equivalence Attention) layer. Shadow embedding acts as a universal technique for converting string-like quantum measurement data into a format suitable for machine learning models. The PEA layer ensures the preservation of permutation invariance in measurements through the use of aggregation functions; 2) We propose SSL4Q, a pioneering semi-supervised learning model specifically designed for quantum data analysis. SSL4Q addresses a critical gap in current quantum computing research by offering a framework for processing quantum data with limited labeled information, a scenario frequently encountered in practical quantum computing applications; 3) We validate the effectiveness of SSL4Q through extensive experiments on quantum state classification using the Heisenberg model and the Variational Quantum Circuit (VQC) model with up to 50 qubits. Despite the sparse labeling of the dataset, SSL4Q demonstrates significant improvements over traditional supervised learning models.

Remark. In our initial investigation, we discovered that Zhang et al. (2023) asserts the classification of quantum states through semi-supervised learning. However, we contend that their method falls short of providing a viable approach for applying machine learning to quantum physics. Their technique, which involves using quantum states directly as inputs, is largely impractical due to the foundational requirement in learning of quantum states: the necessity to first gather quantum data via quantum measurements on the states themselves. Consequently, our method stands as the first to apply semi-supervised learning for the quantum data, addressing the oversight in Zhang et al. (2023).

2. Preliminaries and Related Work

We provide definitions and notations related to quantum computing and refer readers to Nielsen & Chuang (2010) for details. For related work, please see Appendix A.

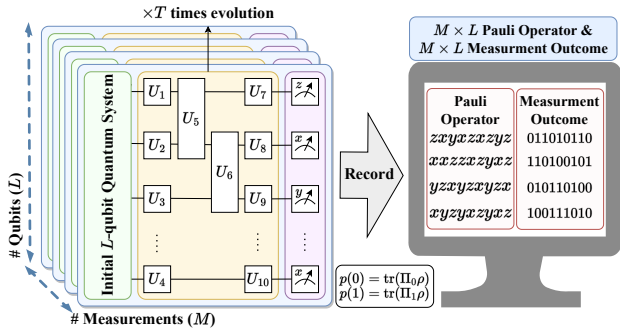

Figure 2. **Generating Quantum Dataset.** The L-qubit quantum system is evolved over T, governed by the Hamiltonian or specified quantum circuits. Random measurements are repeated M times on identical copies of the quantum states and the measurement operators and their outcomes are stored in classical memory.

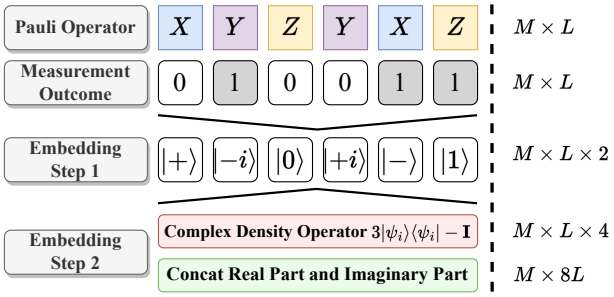

Figure 3. **Shadow Embedding.** The specified measurement operator and its binary output are first transformed into eigenvector representations. Then the density operator for each single-qubit measurement is reconstructed. The final embeddings are obtained by combining the real and imaginary parts of the density matrix.

Qubit and Multi-Qubit Quantum State. *Qubit* is the basic unit in quantum systems. In this paper we consider the spin-1/2 system in which one qubit can be described mathematically as a 2-dimensional normalized complex vector $|\phi\rangle = \left[\begin{smallmatrix}\alpha\\ \beta\end{smallmatrix}\right]$ in the space $\mathbb{C}^2$, where $\alpha, \beta \in \mathbb{C}$ is the *amplitudes* satisfying $|\alpha|^2 + |\beta|^2 = 1$. Such quantum state can be also written as a linear combination $|\phi\rangle = \alpha|0\rangle + \beta|1\rangle$, where $|0\rangle = \left[\begin{smallmatrix}1\\ 0\end{smallmatrix}\right]$ and $|1\rangle = \left[\begin{smallmatrix}0\\ 1\end{smallmatrix}\right]$ are the *Z-basis*, i.e., the eigenvectors of $\boldsymbol{\sigma}^z$ namely the Pauli-Z operator written as $\left[\begin{smallmatrix}1 & 0\\ 0 & -1\end{smallmatrix}\right]$. The same qubit can be decomposed to different orthonormal basis, such as *X-basis* and *Y-basis*, denoted as $|\pm\rangle = \frac{1}{\sqrt{2}}\left[\begin{smallmatrix}1\\ \pm 1\end{smallmatrix}\right]$ and $|\boldsymbol{i}_\pm\rangle = \frac{1}{\sqrt{2}}\left[\begin{smallmatrix}1\\ \pm \boldsymbol{i}\end{smallmatrix}\right]$ respectively. Notation $\boldsymbol{i}$ is the imaginary unit. An alternate formulation for representing a qubit is possible using a tool known as the *density operator* or *density matrix*, e.g., the density matrix of $|0\rangle$ is $\rho_0 = |0\rangle\langle 0| = \left[\begin{smallmatrix}1 & 0\\ 0 & 0\end{smallmatrix}\right]$. See Appendix B for more details.

Projective Measurement. *Quantum measurement* is the testing or manipulation of a physical system to yield a numerical result. A quantum measurement is described by a set of *measurement operators* $\{\boldsymbol{\Pi}_k\}_{k=0}^{K-1}$ satisfying $\sum_k \boldsymbol{\Pi}_k = \mathbf{I}$, where K is the total number of measurement operators. In this paper, we consider the *projective measurement* which means $K = 2$ and $\boldsymbol{\Pi}_k\boldsymbol{\Pi}_{k'} = \delta_{kk'}\boldsymbol{\Pi}_k$. It is easy to check that the density operator of Z-basis, i.e. $\{|0\rangle\langle 0|, |1\rangle\langle 1|\}$ is one of the projective measurements, so are the X-basis and Y-basis. Measuring a qubit leads to collapse of the wave function and produces an outcome k with the probability $p(k)$ satisfying the Born rule (Born, 1926), which states that $p(m) = \mathrm{tr}(\rho\boldsymbol{\Pi}_k)$. For a quantum state with L qubits, performing quantum measurement independently on L qubits is easy to be implemented. The most common strategy is to combine L single-qubit measurement operators to $\boldsymbol{\Pi}_{k,1} \otimes \cdots \otimes \boldsymbol{\Pi}_{k,L}$ (Leibfried et al., 1996; Jullien et al., 2014) where $\otimes$ is the Kronecker product. Such measurement procedure outputs a measurement string $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_L)$ where $\sigma_i \in \{1, \ldots, K\}$.

Classical Shadow. The classical shadow formalism uses randomized (single-shot) measurements to predict many properties of an unknown quantum state, such as phases of quantum matter, correlation function and entanglement entropy of ground state (Huang et al., 2020). Supposing that the quantum state ρ contains L qubits, one can perform randomized single-qubit Pauli measurements and repeat the following procedure a total of M times: (i) prepare an independent copy of the quantum state; (ii) select L single qubit protective measurement uniformly at random (X, Y, Z-basis occur with probability 1/3 each) and (iii) perform the associated measurement to obtain L classical bits. Subsequently, the total $M \times L$ measurement outcomes $\mathbf{O} \in \{0, 1\}^{M\times L}$ are stored in the classical memory. Relevantly, the m-th measurement causes the quantum state to collapse to $S_M = |s_1^{(m)}\rangle \otimes \cdots \otimes |s_L^{(m)}\rangle$ with $|s_l^{(m)}\rangle \in \{|0\rangle, |1\rangle, |+\rangle, |-\rangle, |\boldsymbol{i}_+\rangle, |\boldsymbol{i}_-\rangle\}$ for $l \in \{1, \ldots, L\}$ and $m \in \{1, \ldots, M\}$. Such measurement procedure is simulable in physical quantum computers (Zhang et al., 2021b). Then we can approximate ρ via taking the expectation value of the collection results given as

$$\hat{\rho} = \frac{1}{M}\sum_{t=1}^{M}(3|s_1^{(m)}\rangle\langle s_1^{(m)}| - \mathbf{I}) \otimes \cdots \otimes (3|s_L^{(m)}\rangle\langle s_L^{(m)}| - \mathbf{I}). \tag{1}$$

For predicting properties of quantum states, classical shadow has been proven to be effective for extracting meaningful features from random measurements (Huang et al., 2022; Lewis et al., 2023).

3. Methodology

We first formulate our task in Sec. 3.1. Sec. 3.2 delves into the intricacies of assembling the quantum data and the process of embedding quantum measurement results into a format amenable to learning algorithms. We present the key methods and insights behind SSL4Q in Sec. 3.3.

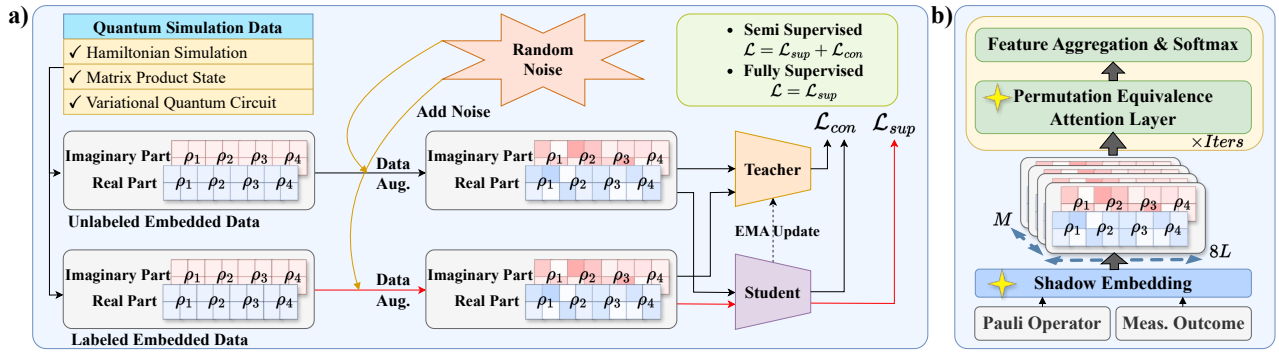

Figure 4. **Pipeline of the Proposed SSL4Q. a)** SSL4Q utilizes semi-supervised learning for quantum data from physical or simulated sources, with partial label masking. Measurement records are encoded through the shadow embedding. The embeddings are augmented randomly before they are fed to the model. SSL4Q incorporates teacher and student models with identical structures but different initializations. The student model learns via standard gradient descent, while the teacher model is an exponential moving average (EMA) of the student model's parameters. **b)** The architecture of the student and the teacher model.

3.1. Problem Formulation

In typical learning-based quantum many body problems, given a set of quantum data $\mathcal{D}^{tr} = \{\mathbf{X}_i, y_i\}_{i=1}^{N^{tr}}$ for training where $\mathbf{X}_i$ generally denotes the embedded measurement results and y denotes its ground-truth prediction target related to certain quantum properties, our objective is to train a model $f_{\boldsymbol{\theta}}$: $\mathbf{X} \rightarrow y$ to predict the labels on the testing dataset $\mathcal{D}^{te} = \{\mathbf{X}_i\}_{i=1}^{N^{te}}$. Throughout this paper, we focus on quantum state classification problem. In our context, the input $\mathbf{X} \in \mathbb{R}^{M \times 8L}$ is conceptualized as shadow embedding while the target y is a one-hot label delineating the category of quantum states with specific properties such as the evolution time and coupling strengths. The model is optimized by the training dataset and identifies categories of unknown quantum systems. Here, M is the number of measurements and L is the number of qubits.

3.2. Quantum Dataset Building from Quantum System

We briefly describe the quantum dataset's definition and its construction methodology. See Sec. 4 for an exhaustive explanation. Consider a quantum system transitioning from an initial state $|\Phi_0\rangle$ to a final state $|\Phi_t\rangle = e^{-\boldsymbol{i}H(w)t}|\Phi_0\rangle$ governed by the Hamiltonian $H(w)$ over time t, with w as the parameters within H. For each sample, we adjust the values of w to generate N^{tr} (resp. N^{te}) distinct final states for the training (resp. testing) dataset. As shown in Fig. 2, for each sample we perform M independent random Pauli measurements on the final state and afterwards store the Pauli operators as well as the measurement outcomes into two separate matrix-like records. Echoing the discussions in Sec. 2, for each measurement on a single copy of $|\Phi_t\rangle$, L measurement operators $\{\{\Pi_{k,l}\}_{k=0}^{1}\}_{l=1}^{L}$ are randomly and independently sampled, where $\{\Pi_{k,l}\} \in \{\{|0\rangle\langle 0|, |1\rangle\langle 1|\}, \{|+\rangle\langle +|, |-\rangle\langle -|\}, \{|\boldsymbol{i}_+\rangle\langle \boldsymbol{i}_+|, |\boldsymbol{i}_-\rangle\langle \boldsymbol{i}_-|\}\}$ corresponding to the eigenvectors of the Pauli-Z, Pauli-X and Pauli-Y operator, respectively. Measuring the l-th qubit with the sampled operator $\{\Pi_{k,l}\}_{k=0}^{1}$ leads to either an output 0 with probability $\mathrm{tr}(\rho_l \Pi_{0,l})$ or 1 with probability $\mathrm{tr}(\rho_l \Pi_{1,l})$. The combined L-qubit measurement operator is performed on $|\Phi_t\rangle$ and we obtain the L-length binary output. The same procedure is repeated M times and we finally have a collection of two $M \times L$ matrix-like records, where one contains the relative Pauli operators and another contains binary measurement outcomes.

However, given that one data format is string-like and the other is binary, concurrently processing these two distinct types of records poses a challenge for machine learning methods. The difficulty is compounded when Pauli strings are converted into matrices, as these contain complex numbers, which are not readily manageable by neural networks designed to operate on real parameters. In Sec. 3.3, we will introduce our approach to embed these disparate types of raw data into a format that is amenable to machine learning. We will also present an effective strategy for learning from such unstructured measurement data.

3.3. SSL4Q

Fig. 4a depicts the architecture of our SSL4Q. it comprises a student and a teacher model, both sharing identical structural designs. The student's learning process is guided by a dual-loss approach, involving supervised loss applied to labeled quantum data and a consistency loss implemented across the entire training dataset. Meanwhile, the teacher is maintained as an exponential moving average (EMA) of the student. Within the SSL4Q framework, we propose two critical innovations crafted for unordered quantum measurement data: the shadow embedding and the permutation equivalence attention (PEA) layer.

3.3.1. SHADOW EMBEDDING

We develop a simple yet effective method for embedding Pauli operators and their measurement outcomes, drawing inspiration from the concept of classical shadow (Huang et al., 2020). For simplicity, below we discuss the shadow embedding for a single input sample. As depicted in Fig. 3, the initial step involves producing the corresponding Pauli basis, derived from the Pauli operator and binary outcome (Step 1). Following this, we generate the density matrix representation for each outcome, expressed as $\rho_i = 3|\psi_i\rangle\langle\psi_i| - \mathbf{I}$ for $i \in \{1, \ldots, L\}$, where $|\psi_i\rangle$ denotes the i-th basis state determined by the sampled Pauli operator and the binary outcome, resulting in a complex-valued matrix. Subsequently, rather than performing a Kronecker product of all the L single-qubit density matrices as the operations in Eq. 1, which would result in an impractically large size relative to the system, we adopt a more efficient approach. This involves concatenating the real and imaginary parts of each 2×2 complex density operator into an 8-element vector (Step 2), transforming two challenging $M \times L$ data formats into a manageable $M \times 8L$ matrix, designated as the input $\mathbf{X}$.

It is important to note that previous research has provided insights into the definition and processing methodologies of quantum data. However, unlike studies such as Xiao et al. (2022); Tang et al. (2024), which primarily rely on discrete measurement outputs, our model uniquely incorporates information from random measurement operators, with the hope of capturing a more comprehensive representation of the quantum state. Contrasting with methods that average all measurement results, as seen in Zhu et al. (2022); Du et al. (2023), we treat each measurement as an individual entity, with the hope of retaining the essential information inherent in the measurement data. Also, our model contains a PEA layer with aggregation function, specifically designed for this type of data. This layer is distinctly advantageous compared to the approach in (Qian et al., 2023), which utilizes PointNet (Qi et al., 2017) to process shadow-like inputs.

3.3.2. PERMUTATION EQUIVALENCE ATTENTION LAYER

Contrasting with the recent quantum systems learning model (Tang et al., 2024), which treats quantum data as ordered and language-like inputs, SSL4Q acknowledges and preserves the inherently *unordered* nature of random measurements. Specifically, for a shadow embedding $\mathbf{X} \in \mathbb{R}^{M \times 8L}$, our aim is to develop a permutation-invariant neural network structure $f_{\boldsymbol{\theta}}$ for both the student and teacher models. This network satisfies the condition $f_{\boldsymbol{\theta}}(\pi_1 \circ \mathbf{X}) = \pi_1 \circ f_{\boldsymbol{\theta}}(\mathbf{X})$ for any permutation π_1, where π_i represents the permutation along the i-th axis. The permutation invariance is usually realized by a network that performs pooling over embeddings extracted from the permutation equivalence layer (Wu et al., 2020). We incorporate a permutation equivalence attention (multi-)layer (PEA) followed by a feed-forward network layer (FFN) to effectively discern patterns from unordered quantum data. An aggregation operation is integrated into the final layer to ensure the entire model's permutation invariance.

As seen in Fig. 4b, the core of both student and teacher models is akin to the multi-layer transformer decoder (Vaswani et al., 2017). The distinction lies in our omission of positional encoding and exclusion of dropout within the attention layer. Suppose that the $(f-1)$-th layer's output $\mathbf{H}^{f-1} \in \mathbb{R}^{M \times d}$ where d is the hidden dimension, we have:

$$\begin{aligned}
\overline{\mathbf{H}}^f &= \text{PEA}(\mathbf{H}^{f-1}) = [\mathbf{O}_1^f, \ldots, \mathbf{O}_h^f]\mathbf{W}_{o_1}^f, \\
\mathbf{H}^f &= \text{FFN}(\overline{\mathbf{H}}^f) = \overline{\mathbf{H}}^f \mathbf{W}_{o_2}^f, \\
\text{with } \mathbf{O}_j^f &= \frac{(\mathbf{H}^{f-1}\mathbf{W}_{q,j}^f)(\mathbf{H}^{f-1}\mathbf{W}_{k,j}^f)^\top}{\sqrt{d}} \mathbf{H}^{f-1}\mathbf{W}_{v,j}^f,
\end{aligned} \tag{2}$$

where $[\cdot]$ denotes the concatenation operation, $\mathbf{W}_{q,j}^f$, $\mathbf{W}_{k,j}^f$, $\mathbf{W}_{v,j}^f \in \mathbb{R}^{d \times d/h}$ and $\mathbf{W}_{o_1}^f, \mathbf{W}_{o_2}^f \in \mathbb{R}^{d \times d}$ are the parameters to be learned. The same process is looped F times and then we use the average pooling operation to turn the output $\mathbf{H}^F$ of the last layer into a d-dimensional feature vector.

3.3.3. TRAINING THE SSL4Q

Having outlined the key components of SSL4Q, we now turn our focus to its training methodology. The student and teacher models, while structurally identical, commence with distinct random parameter initializations. In the rest of the paper, we refer to the student model as $f_{\boldsymbol{\theta}}$ with parameters $\boldsymbol{\theta}$, and teacher model as $f'_{\boldsymbol{\theta}'}$ with $\boldsymbol{\theta}'$. In each training iteration, a combination of labeled and unlabeled shadow embeddings is randomly selected. Allocating a fraction of each batch to labeled examples is beneficial, ensuring that the supervised training signal is sufficiently robust from the outset to enable rapid learning and avoid the pitfalls of uncertainty. In our experiments, the proportion of labeled samples is set to between $1/8$ and $1/4$ of the total batch size. Prior to processing the shadow embeddings, each sample is subjected to data augmentation, a step crucial for the computation of the consistency loss. The extent of augmentation can differ between labeled and unlabeled samples. We opt for Gaussian noise since we find that it significantly enhances model performance despite its simplicity. The student model learns through both supervised loss and consistency loss $\mathcal{L} = \mathcal{L}_{sup} + \lambda \mathcal{L}_{con}$, where

$$\begin{aligned}
\mathcal{L}_{sup} &= -\frac{1}{B} \sum_{i \in \mathcal{B}} \sum_{c=1}^{C} \mathbb{I}\left[y_i = c\right] \log\left(f_{\boldsymbol{\theta}}\left(\mathbf{X}_i\right)_c\right), \\
\mathcal{L}_{con} &= \frac{1}{B} \sum_{i \in \mathcal{B}} \left| f_{\boldsymbol{\theta}}(\mathbf{X}_i) - f'_{\boldsymbol{\theta}'}(\mathbf{X}_i) \right|^2,
\end{aligned} \tag{3}$$

Table 1. Classification accuracy in identifying different classes of Heisenberg model quantum systems w.r.t. different number of measurements M. The mask rate is fixed as $\eta = 0.9$. The SSL4Q sup. denotes the variant of SSL4Q trained only by labeled samples using supervised loss. The best results are emphasized in red while the second-best results are distinguished in blue.

Method	$M=16$				$M=32$				$M=64$				$M=128$			
	$L=20$	$L=30$	$L=40$	$L=50$	$L=20$	$L=30$	$L=40$	$L=50$	$L=20$	$L=30$	$L=40$	$L=50$	L=20	$L=30$	$L=40$	$L=50$
Gaussian Kernel	53.73	54.87	55.07	60.90	56.70	57.87	61.80	56.30	59.33	62.37	59.30	57.83	66.57	60.67	49.37	59.30
NTK	53.17	54.27	55.03	60.67	56.17	57.27	61.07	55.07	59.47	61.03	59.33	57.00	66.30	60.17	49.30	58.80
RNN	61.17	64.47	62.90	66.13	72.60	70.93	69.47	69.67	81.93	74.53	72.50	73.13	85.83	83.43	76.90	73.40
LLM4QPE	63.90	64.13	64.53	67.60	76.73	73.80	70.23	**70.57**	83.20	**77.30**	74.83	74.57	88.70	84.87	79.13	75.57
SSL4Q (sup.)	**64.07**	**65.20**	**65.47**	**68.07**	**78.90**	**75.87**	**72.77**	70.23	**83.67**	76.27	**75.73**	**74.90**	**92.77**	**86.23**	**82.17**	**81.67**
SSL4Q	**71.97**	**68.63**	**72.87**	**71.40**	**84.53**	**80.13**	**78.03**	**73.40**	**89.77**	**81.50**	**82.43**	**83.60**	**93.83**	**88.00**	**83.83**	**84.43**

where λ is the consistency weight, B denotes the batch size, C is the number of classes and $\mathbb{I}[\cdot]$ is an indicator function. Whereas the teacher model's parameters are an EMA of the student's, i.e., $\boldsymbol{\theta}'_u = \alpha\boldsymbol{\theta}'_{u-1} + (1-\alpha)\boldsymbol{\theta}_u$ for the u-th iteration where α is a smoothing coefficient. Our model can also be easily adapted for fully supervised learning mode which entails using only labeled shadow embeddings in each iteration and omitting data augmentation. The numerical experiments reveal that incorporating Gaussian noise and consistency loss into the semi-supervised learning framework significantly enhances SSL4Q's efficacy in classification of quantum states.

4. Experiments

We perform experiments on two types of quantum models: the Heisenberg model (Takahashi & Suzuki, 1972), which is important for Hamiltonian simulation (Feynman et al., 2018) in quantum information sciences, and the VQC model (Cerezo et al., 2021) noted for its applicable to quantum algorithms for near-term quantum devices (Schuld et al., 2015). We utilize the VQC to encode classical inputs from the MNIST dataset (LeCun et al., 1998) into quantum states through parameterized quantum gates. For both models, we experiment with different mask rates η, randomly obscuring a portion of the training data labels.

Baselines. To our best knowledge, there is no semi-supervised models for quantum many body problems. Thus we only consider fully supervised baselines. For the Heisenberg model, we establish our baselines using kernel methods including the Gaussian Kernel and Neural Tangent Kernel (NTK) implemented in Huang et al. (2022). To broaden our comparative scope, we also explore advanced deep learning approaches. These include a Recurrent Neural Network (RNN)-based model (Carrasquilla et al., 2019), and a transformer-based model LLM4QPE (Tang et al., 2024). For our experiments involving the VQC model, we consider the convolutional neural network (CNN) (He et al., 2016) and the quantum CNN (QCNN) (Cong et al., 2019). In both cases, we utilize the implementation provided by Hur et al. (2022) for comparison. In this paper, all our experimental results are reported for a model configuration consisting of 4 heads, 2 layers, and 128 hidden dimensions. We refer the readers to Appendix D for more about the hyper-parameter configurations of SSL4Q and baselines. To demonstrate the effectiveness of the proposed semi-supervised learning framework, a fully supervised variant (SSL4Q sup.) is introduced to evaluate the model's ability to effectively utilize the abundance of unlabeled data for model's regularization.

4.1. Experiments on Heisenberg Model

Our first investigation focuses on the quantum system undergoing time evolution governed by a parameterized Hamiltonian. We start by introducing the fundamentals of generating the datasets and then we report the numerical results.

4.1.1. Dataset

We examine the one-dimensional transverse-field Ising model, described by the Hamiltonian:

$$H(J,g) = -J\left(\sum_{\langle ij\rangle} \boldsymbol{\sigma}_i^z \boldsymbol{\sigma}_j^z + g\sum_i \boldsymbol{\sigma}_i^x\right), \qquad (4)$$

where $\langle ij\rangle$ represents adjacent qubits i and j, and $\boldsymbol{\sigma}_i^z$ and $\boldsymbol{\sigma}_i^x$ denote the Pauli-Z and Pauli-X operator acting on the i-th qubit, respectively. For specific values of J and g, the quantum system evolves over time as $|\Phi_t\rangle = e^{-\boldsymbol{i}Ht}|\Phi_0\rangle$. We explore various evolution times $t \in \{5, 10, 20\}$ms. Different time evolution leads to a distinguishable variation pattern such as entanglement entropy (Xiao et al., 2022), resulting in three classes of quantum systems based on time t. Additionally, the dynamic evolution of the quantum system is influenced by the strength of J and g. We set $J = 1$ and assume g is uniformly distributed within $[0, 2]$. The system enters an ordered phase for $|g| < 1$ and an unordered phase for $|g| > 1$ (Heyl et al., 2013), leading to two classes based on phase. Consequently, there are a total of 6 distinct classes, each characterized by its own phase and evolution time. For the testing dataset, we randomly select 1000 different values of g for each $t \in \{5, 10, 20\}$ms, resulting in $N^{te} = 3000$. For the training dataset, we randomly sample the time step t and the strength g to generate data and we set $N^{tr} = 2000$. Simulating the quantum system with up to

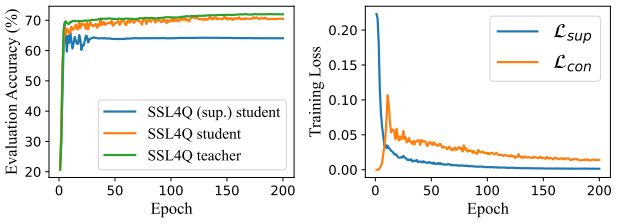

Figure 5. **Left)** Classification Accuracy of teacher model and student model on the testing dataset. **Right)** The curve of $\mathcal{L}_{sup}$ and $\mathcal{L}_{con}$ of SSL4Q on the training dataset. For illustrative clarity, this visualization is confined to the initial 200 training epochs.

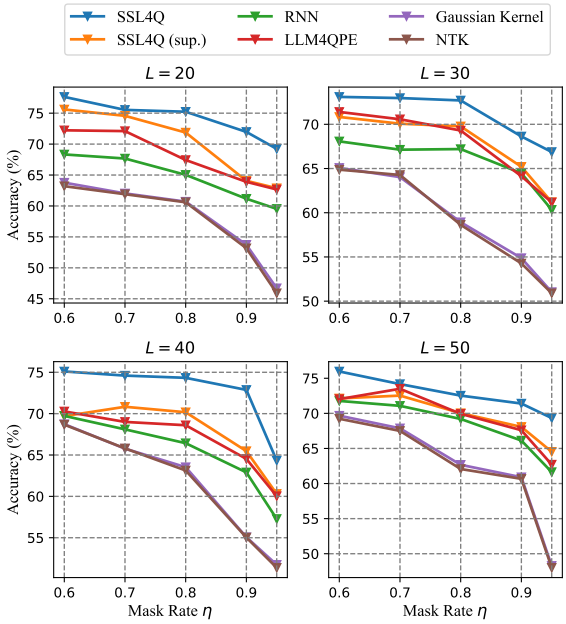

Figure 6. Classification accuracy for Heisenberg model quantum systems by mask rates η. The number of measurements is M=16.

50 qubits is generally nontrivial and we employ the matrix product state (MPS) (Orús, 2014) to represent the quantum state and generate the measurement records. More details can be found in Appendix C. We vary the mask rate η and the number of measurements M for different system sizes and examine the performance among SSL4Q and baselines.

4.1.2. Experimental Results

We perform experiments on various system sizes $L \in \{20, 30, 40, 50\}$, with constant mask rate $\eta = 0.9$, implying that merely 10% (200 out of 2000) of the training samples are labeled. Our focus is on the most challenging scenario with M=16, where a lower count of measurements introduces greater uncertainties in the resultant shadow embeddings. As listed in Tab. 1, SSL4Q consistently outperforms baselines, achieving the highest mean accuracy across all system sizes. Intriguingly, even when the parameter updates rely solely on the supervised loss from the labeled samples, it consistently outperforms baseline models, including LLM4QPE in which transformer is used akin to ours. A notable distinction in SSL4Q is the absence of a LSTM layer for embedding, which in LLM4QPE transforms discrete measurement outcomes into continuous features, thereby disrupting the permutation invariance characteristic of quantum measurements. Moreover, the number of LSTM's parameters linearly scales with the number of measurements M, predisposing the model to overfit training data yet underfitting in test.

We further investigate the efficiency of training SSL4Q with and without supervised labels. As shown in Fig. 5 (left), the EMA-weighted model (i.e., the teacher model) provides more accurate predictions than the bare student models after an initial period. For the SSL4Q variant that relies exclusively on supervised loss, we report the student model's evaluation accuracy, as the teacher model's consistency loss is deactivated. In contrast, in the generic SSL4Q framework, the teacher model's evaluation accuracy is emphasized due to its superior performance and generalization capabilities on test datasets. This performance distinction is illustrated in Fig. 5 (right). In the early training phase (before the 20th epoch), the model rapidly adapts to the labeled training samples, evidenced by the precipitous decline in supervised loss. This supervised training signal facilitates rapid training and prevents entrapment in uncertainty. For comparison, the consistency loss escalates sharply at first, as we adopt a ramp-up schedule for the consistency loss coefficient λ to stabilize the training (Tarvainen & Valpola, 2017). As training progresses (after the 20th epoch), the consistency loss assumes a dominant role in the loss function, thereby increasing the penalty for output inconsistencies. This strategic shift allows the model to progressively enhance its adaptation to the inherent uncertainties of quantum measurements.

In Fig. 6, we study the impact of varying mask levels on the model's performance. A higher mask rate indicates fewer labeled training samples, introduce an augmented challenge to the model's capacity for learning and generalization within datasets constrained by limited labels. It's crucial to note that all baseline models rely solely on supervised loss for parameter updates, fundamentally due to their fully supervised nature and a lack of consideration for semi-supervised learning frameworks by their original developers. To this end, a supervised learning variant of SSL4Q (SSL4Q sup.) is designed to evaluate the SSL4Q's ability to effectively utilize the abundance of unlabeled data for model's regularization. For a fair comparison, a shadow embedding layer is utilized as the input across all models. SSL4Q exhibits consistent superiority over the baselines, with an improvement margin ranging between 2% and 5% except in the cases of $L = 20$ and $L = 50$ with $\eta = 0.7$. Notably, SSL4Q (sup.) also marginally exceeds the other two neural network-based

methods including RNN and LLM4QPE. This advantage is ascribed to our innovative PEA layer. Moreover, neural network-based methods generally surpass kernel methods, aligning with the results reported in papers Wang et al. (2022); Tang et al. (2024), thus underscoring the superiority of deep learning approaches over traditional kernel-based machine learning methods in representing quantum states and their characteristics.

4.2. Experiments on VQC

We then delve into the Variational Quantum Circuit (VQC), an approach in harnessing quantum advantage on Noisy Intermediate-Scale Quantum (NISQ) devices (Cerezo et al., 2021). Below we briefly discuss how we employ VQC to construct the quantum dataset.

4.2.1. Dataset

Given a set of classical data $\boldsymbol{\omega}_r \in \mathbb{R}^P$ for $r \in \{1, \ldots, R\}$, VQC can be envisioned as a quantum feature map encoding classical data into quantum states:

$$\mathbf{U}(\boldsymbol{\omega}) = \mathbf{U}_R(\boldsymbol{\omega}_R) \cdots \mathbf{U}_2(\boldsymbol{\omega}_2)\mathbf{U}_1(\boldsymbol{\omega}_1),$$
$$\text{with } \mathbf{U}_r(\boldsymbol{\omega}_r) = \prod_p e^{-i\boldsymbol{\omega}_{r,p}H_{r,p}/2}, \tag{5}$$

where $\mathbf{U}_r(\boldsymbol{\omega}_r)$ is a unitary satisfying $\mathbf{U}_r(\boldsymbol{\omega}_r)\mathbf{U}_r^\dagger(\boldsymbol{\omega}_r) = \mathbf{I}$ ($\dagger$ denotes the conjugate transpose), $H_{r,p}$ is a Hermitian operator satisfying $H_{r,p}^\dagger = H_{r,p}$ and $\boldsymbol{\omega}_{r,p}$ is the p-th element in $\boldsymbol{\omega}_r$. We follow the pre-processing protocal of Hur et al. (2022), wherein each 28×28 image is condensed to a 16-dimensional feature vector $\boldsymbol{v}$ via Principal Component Analysis (PCA). We set $R = 2$ and $P = 8$ to encode the image into its quantum representation, where $\boldsymbol{v}_{[1:8]}$ is assigned to $\boldsymbol{\omega}_1$ and $\boldsymbol{v}_{[9:16]}$ to $\boldsymbol{\omega}_2$. The Hermitian operator is chosen to be $H_{1,p} = \boldsymbol{\sigma}_x$ and $H_{2,p} = \boldsymbol{\sigma}_y$ for $p \in \{1, \ldots, 8\}$. This specific quantum circuit is applied to the initial state to produce an 8-qubit quantum representation of an image, denoted as $|\Psi(\boldsymbol{\omega})\rangle = \mathbf{U}(\boldsymbol{\omega})|0\rangle^{\otimes 8}$. Subsequent measurements from this quantum state yield the necessary measurement records, with further details available in Appendix C. The image's category is treated as the class identifier for the quantum systems. In this paper we consider binary classification. We generate a quantum dataset including images of class 0 and class 1, resulting in a training dataset of size $N^{tr} = 12665$ and a testing dataset of size $N^{te} = 2115$.

4.2.2. Experimental Results

The average classification accuracy across three runs with distinct random seeds is listed in Tab. 2. We incrementally adjust the mask rate η from 0.6 to 0.95 while maintaining $M = 512$ and $N^{tr} = 12665$. Our SSL4Q framework, along with its supervised variant SSL4Q (sup.), outperforms the other two baseline models across all evaluated mask rates.

Table 2. Accuracy in predicting categories of quantum states encoded by '0' and '1' of MNIST. We fix $M = 512$ and $N^{tr} = 12665$. Best and second-best are in red and blue, respectively.

Method	$\eta = 0.6$	$\eta = 0.7$	$\eta = 0.8$	$\eta = 0.9$	$\eta = 0.95$
CNN	88.82	92.85	94.50	86.10	93.24
QCNN	97.40	96.80	96.95	96.38	95.74
SSL4Q (sup.)	**98.44**	**98.77**	**99.01**	**98.68**	**98.96**
SSL4Q	**99.20**	**99.15**	**99.20**	**99.10**	**99.24**

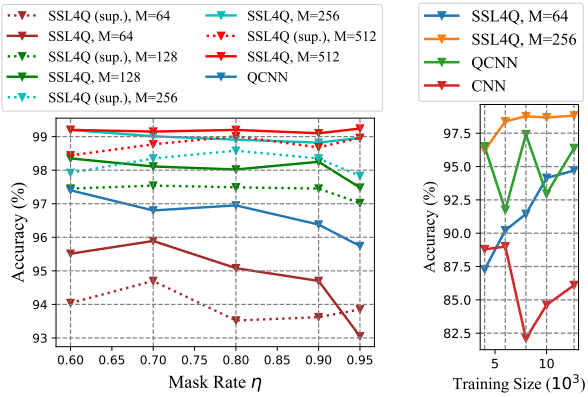

Figure 7. **Left)** Classification accuracy for VQC models across various mask rates with $N^{tr} = 12665$. **Right)** Accuracy among different sizes of training dataset with $\eta = 0.9$.

Notably, SSL4Q achieves exceptional accuracy, consistently exceeding 90% even at a mask rate of 0.95.

Further study into how the number of measurements affects model performance is shown in Fig. 7 (Left), SSL4Q and SSL4Q (sup.) consistently surpass QCNN when $M \geq 128$, with performance improving as M increases. This suggests that directly applying VQC as quantum neural networks for processing classical data does not fully leverage the benefits of quantum computing, often yielding results inferior to those of well-engineered classical neural networks (Cerezo et al., 2021). Although information may be lost when performing limited measurements, the results show that SSL4Q can effectively discern the quantum systems corresponding to image categories from the measurement records. Additionally, an analysis of the impact of various training set sizes on model performance is presented in Fig. 7 (Right). It is observed that SSL4Q's prediction accuracy at $M = 256$ surpasses both QCNN and CNN, significantly improving upon SSL4Q's accuracy at $M = 64$. With increasing training set size, SSL4Q's prediction accuracy exhibits a progressively upward trend, with notably reduced oscillation amplitude compared to QCNN and CNN.

Table 3. Comparison of classification accuracy for VQC model under noisy conditions with different depolarizing noise levels.

Method	$p = 0.01$	$p = 0.05$	$p = 0.1$
SSL4Q	98.71	98.34	97.68
QCNN	94.47	92.59	72.34

4.3. Further Discussions

Ablation on Shadow Embedding and PEA Layer. To evaluate the efficacy of shadow embedding, we substitute it with discrete measurement outcomes. We also replace the PEA layer with an LSTM layer. The results in Tab. 4 located Appendix E demonstrate that incorporating information from the measurement operators can significantly enhance the model's performance. Moreover, compared to the LSTM layer that overlooks the unordered nature of quantum measurements, SSL4Q's adoption of permutation invariance further boosts its performance.

Impact of Different Levels of Noise. We simulate gate noise and measurement noise as reported in IBM quantum computers and simulators, and collect noisy quantum datasets using the VQC model. Then the model is re-trained from scratch to assess model's performance under noisy conditions. For a comparison, we present the accuracy of SSL4Q in predicting categories of quantum states encoded by '0' and '1' of MNIST under *noiseless* conditions. This will serve as a benchmark for comparison with the model's performance under noisy conditions. The experiments are repeated 5 times with different random seed initailizations and averaged classification accuracy of SSL4Q is 98.82%, consistently outperforming CNN's 86.1% and QCNN's 96.38%. We first evaluate the model's performance under a common type of gate noise, specifically depolarizing noise. The output state of the input state ρ after the noise channel is given as $\mathcal{E}(\rho) = \sum_i K_i \rho K_i^\dagger$, where $\{K_i\}$ are the Kraus operators satisfying $\sum_i K_i K_i^\dagger = I$. Concretely, these operators are written as

$$K_0 = \sqrt{1-p}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \; K_1 = \sqrt{p/3}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$
$$K_2 = \sqrt{p/3}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \; K_3 = \sqrt{p/3}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{6}$$

The depolarizing noise model is extracted from the `simulator_statevector` simulator from IBM quantum cloud service, setting the intensity of depolarizing noise p at three levels: 0.01, 0.05, and 0.1. The accuracy of SSL4Q in predicting categories of quantum states under the gate noisy conditions are listed in Tab. 3. It is observed that while accuracy decreases with the increase in noise intensity, the model's accuracy remains higher than that of the baselines including CNN and QCNN under noiseless conditions, even at a noise intensity of $p = 0.1$. Furthermore, we assess the SSL4Q's performance under measurement error. Specifically, we reference the reported measurement errors of the IBM quantum computer `ibm_kyiv`, setting `meas0_prev1` as 0.0088 (the probability that the measurement result of $|1\rangle$ is that of $|0\rangle$) and `meas1_prev0` as 0.0042 (the probability that the measurement result $|0\rangle$ is that of $|1\rangle$). The experimental results show an accuracy in predicting categories of quantum states of 98.91%. This result also surpasses the baselines CNN and QCNN under noiseless conditions.

Impact of Measurement Uncertainties. We construct a toy dataset for the Heisenberg model specified by $t = 5$ms, $g = 1$, $L = 50$, $\eta = 0.9$, and $M = 16$, with varied $g \in \{0.5, 1.5\}$. The dataset is constructed with equal size for training and testing ($N^{tr} = N^{te} = 100$), with balanced positive and negative samples. This setup is intended for a binary classification task, aiming to assess predictive accuracy under conditions of limited measurements with large uncertainties. As demonstrated in Fig. 12 in Appendix E, SSL4Q exhibits enhanced robustness in scenarios of high measurement uncertainty. This is attributed to our strategy of preserving the intrinsic unorder of quantum measurements, contrary to approaches that assume ordered embeddings, as seen in RNN and LLM4QPE implementations.

Impact of Data Augmentation and Consistency Weight. We study the effects of the noise used for data augmentation and the consistency weight λ. The results are depicted in Fig. 13. The predictive error (one minus accuracy) is reported on the dataset with $L = 20$, $M = 32$ and $\eta = 0.9$. Our findings indicate a nuanced relationship between test accuracy and the standard deviation (std) of Gaussian noise, where accuracy initially increases before declining as the noise shifts from low to high values, typically peaking at a std of 0.25. Besides, when the λ is relatively small, the test accuracy improves. However, when λ is larger than a threshold (10 under the dataset's settings), the accuracy would drop. In our experiments, we adopt a Gaussian noise setting with a mean of 0 and a std of 0.25, alongside a λ value of 10. This configuration is found to be optimal across a broad range of settings, although further fine-tuning may yield enhanced performance.

5. Conclusion and Outlook

We have introduced a semi-supervised approach, SSL4Q, tailored for quantum data analysis with limited labeling. Our empirical results show its superiority over conventional supervised models in classification of quantum states of the Heisenberg and VQC models, under varying label masking rates and measurement quantities. Future efforts will explore the adaptability of our model to diverse quantum systems.

Impact Statement

Quantum computing has been a promising paradigm for artificial intelligence, with SSL4Q exemplifying its potential through semi-supervised learning in quantum state classification. This approach leverages limited labeled alongside abundant unlabeled data, enhancing the improvements of quantum technologies. While promising for quantum technology advancement, it also poses ethical challenges, such as potential misuse and increased digital divide. We emphasize the importance of responsible research and policy frameworks to mitigate these risks, advocating for an interdisciplinary dialogue to ensure equitable and positive societal contributions from quantum advancements.

Acknowledgement

We would like to thank the anonymous reviewers for their valuable feedback and suggestions that help to improve this work. This work was in part supported by National Natural Science Foundation of China (92370201, 62222607, 72342023).

References

Bergholm, V., Izaac, J., Schuld, M., Gogolin, C., Ahmed, S., Ajith, V., Alam, M. S., Alonso-Linaje, G., Akash-Narayanan, B., Asadi, A., et al. Pennylane: Automatic differentiation of hybrid quantum-classical computations. *arXiv preprint arXiv:1811.04968*, 2018.

Born, M. Quantenmechanik der stoßvorgänge. *Zeitschrift für physik*, 38(11-12):803–827, 1926.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., et al. Language models are few-shot learners. *Advances in neural information processing systems*, 33: 1877–1901, 2020.

Burgarth, D. and Yuasa, K. Quantum system identification. *Physical Review Letters*, 108(8):080502, 2012.

Carleo, G. and Troyer, M. Solving the quantum many-body problem with artificial neural networks. *Science*, 355 (6325):602–606, 2017.

Carleo, G., Cirac, I., Cranmer, K., Daudet, L., Schuld, M., Tishby, N., Vogt-Maranto, L., and Zdeborová, L. Machine learning and the physical sciences. *Reviews of Modern Physics*, 91(4):045002, 2019.

Carrasquilla, J. Machine learning for quantum matter. *Advances in Physics: X*, 5(1):1797528, 2020.

Carrasquilla, J., Torlai, G., Melko, R. G., and Aolita, L. Reconstructing quantum states with generative models. *Nature Machine Intelligence*, 1(3):155–161, 2019.

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. *Nature Reviews Physics*, 3(9):625–644, 2021.

Cha, P., Ginsparg, P., Wu, F., Carrasquilla, J., McMahon, P. L., and Kim, E.-A. Attention-based quantum tomography. *Machine Learning: Science and Technology*, 3(1): 01LT01, 2021.

Chen, Y., Mancini, M., Zhu, X., and Akata, Z. Semi-supervised and unsupervised deep visual learning: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 2022.

Chen, Z., Newhouse, L., Chen, E., Luo, D., and Soljacic, M. ANTN: Bridging autoregressive neural networks and tensor networks for quantum many-body simulation. *Advances in neural information processing systems*, 2023.

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., and Bengio, Y. Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*, 2014.

Cong, I., Choi, S., and Lukin, M. D. Quantum convolutional neural networks. *Nature Physics*, 15(12):1273–1278, 2019.

Corboz, P. Variational optimization with infinite projected entangled-pair states. *Physical Review B*, 94(3):035133, 2016.

D'Ariano, G. M., Paris, M. G., and Sacchi, M. F. Quantum tomography. *Advances in imaging and electron physics*, 128:206–309, 2003.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 1:4171–4186, 2019.

Du, Y., Yang, Y., Liu, T., Lin, Z., Ghanem, B., and Tao, D. Shadownet for data-centric quantum system learning. *arXiv preprint arXiv:2308.11290*, 2023.

Eisert, J., Hangleiter, D., Walk, N., Roth, I., Markham, D., Parekh, R., Chabaud, U., and Kashefi, E. Quantum certification and benchmarking. *Nature Reviews Physics*, 2(7):382–390, 2020.

Feynman, R. P. et al. Simulating physics with computers. *Int. j. Theor. phys*, 21(6/7), 2018.

Fishman, M., White, S. R., and Stoudenmire, E. M. The ITensor Software Library for Tensor Network Calculations. *SciPost Phys. Codebases*, pp. 4, 2022. doi: 10.21468/SciPostPhysCodeb.4. URL https://scipost.org/10.21468/SciPostPhysCodeb.4.

Flammia, S. T., Gross, D., Liu, Y.-K., and Eisert, J. Quantum tomography via compressed sensing: error bounds, sample complexity and efficient estimators. *New Journal of Physics*, 14(9):095022, 2012.

Gammelmark, S. and Mølmer, K. Fisher information and the quantum cramér-rao sensitivity limit of continuous measurements. *Physical review letters*, 112(17):170401, 2014.

Gao, M., Zhang, Z., Yu, G., Arık, S. Ö., Davis, L. S., and Pfister, T. Consistency-based semi-supervised active learning: Towards minimizing labeling cost. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part X 16*, pp. 510–526. Springer, 2020.

Gebhart, V., Santagati, R., Gentile, A. A., Gauger, E. M., Craig, D., Ares, N., Banchi, L., Marquardt, F., Pezzè, L., and Bonato, C. Learning quantum systems. *Nature Reviews Physics*, 5(3):141–156, 2023.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014.

Hatano, N. and Suzuki, M. Finding exponential product formulas of higher orders. In *Quantum annealing and other optimization methods*, pp. 37–68. Springer, 2005.

He, K., Zhang, X., Ren, S., and Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

Heyl, M., Polkovnikov, A., and Kehrein, S. Dynamical quantum phase transitions in the transverse-field ising model. *Physical review letters*, 110(13):135704, 2013.

Hibat-Allah, M., Ganahl, M., Hayward, L. E., Melko, R. G., and Carrasquilla, J. Recurrent neural network wave functions. *Physical Review Research*, 2(2):023358, 2020.

Howard, J. and Ruder, S. Universal language model fine-tuning for text classification. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 328–339. Association for Computational Linguistics, July 2018.

Hu, H.-Y., Choi, S., and You, Y.-Z. Classical shadow tomography with locally scrambled quantum dynamics. *Physical Review Research*, 5(2):023027, 2023.

Huang, H.-Y., Kueng, R., and Preskill, J. Predicting many properties of a quantum system from very few measurements. *Nature Physics*, 16(10):1050–1057, 2020.

Huang, H.-Y., Kueng, R., Torlai, G., Albert, V. V., and Preskill, J. Provably efficient machine learning for quantum many-body problems. *Science*, 377(6613):eabk3333, 2022.

Hur, T., Kim, L., and Park, D. K. Quantum convolutional neural network for classical data classification. *Quantum Machine Intelligence*, 4(1):3, 2022.

Jullien, T., Roulleau, P., Roche, B., Cavanna, A., Jin, Y., and Glattli, D. Quantum tomography of an electron. *Nature*, 514(7524):603–607, 2014.

Karagiorgi, G., Kasieczka, G., Kravitz, S., Nachman, B., and Shih, D. Machine learning in the search for new fundamental physics. *Nature Reviews Physics*, 4(6):399–412, 2022.

Khan, F. H., Qamar, U., and Bashir, S. A semi-supervised approach to sentiment analysis using revised sentiment strength based on sentiwordnet. *Knowledge and information Systems*, 51:851–872, 2017.

Kliesch, M. and Roth, I. Theory of quantum system certification. *PRX quantum*, 2(1):010201, 2021.

LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

Leibfried, D., Meekhof, D., King, B., Monroe, C., Itano, W. M., and Wineland, D. J. Experimental determination of the motional quantum state of a trapped atom. *Physical review letters*, 77(21):4281, 1996.

Lewis, L., Huang, H.-Y., Tran, V. T., Lehner, S., Kueng, R., and Preskill, J. Improved machine learning algorithm for predicting ground state properties. *arXiv preprint arXiv:2301.13169*, 2023.

Liu, Q., Yu, L., Luo, L., Dou, Q., and Heng, P. A. Semi-supervised medical image classification with relation-driven self-ensembling model. *IEEE transactions on medical imaging*, 39(11):3429–3440, 2020a.

Liu, Y., Deng, G., Zeng, X., Wu, S., Yu, Z., and Wong, H.-S. Regularizing discriminative capability of cgans for semi-supervised generative learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 5720–5729, 2020b.

Miles, C., Samajdar, R., Ebadi, S., Wang, T. T., Pichler, H., Sachdev, S., Lukin, M. D., Greiner, M., Weinberger, K. Q., and Kim, E.-A. Machine learning discovery of new phases in programmable quantum simulator snapshots. *Physical Review Research*, 5(1):013026, 2023.

Nielsen, M. A. and Chuang, I. L. *Quantum computation and quantum information*. Cambridge university press, 2010.

Orús, R. A practical introduction to tensor networks: Matrix product states and projected entangled pair states. *Annals of physics*, 349:117–158, 2014.

Orús, R. Tensor networks for complex quantum systems. *Nature Reviews Physics*, 1(9):538–550, 2019.

Peters, M. E., Ammar, W., Bhagavatula, C., and Power, R. Semi-supervised sequence tagging with bidirectional language models. *arXiv preprint arXiv:1705.00108*, 2017.

Qi, C. R., Su, H., Mo, K., and Guibas, L. J. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 652–660, 2017.

Qian, Y., Du, Y., He, Z., Hsieh, M.-h., and Tao, D. Multimodal deep representation learning for quantum cross-platform verification. *arXiv preprint arXiv:2311.03713*, 2023.

Schuld, M., Sinayskiy, I., and Petruccione, F. An introduction to quantum machine learning. *Contemporary Physics*, 56(2):172–185, 2015.

Sharir, O., Levine, Y., Wies, N., Carleo, G., and Shashua, A. Deep autoregressive models for the efficient variational simulation of many-body quantum systems. *Physical review letters*, 124(2):020503, 2020.

Shi, Y.-Y., Duan, L.-M., and Vidal, G. Classical simulation of quantum many-body systems with a tree tensor network. *Physical review a*, 74(2):022320, 2006.

Søgaard, A. *Semi-supervised learning and domain adaptation in natural language processing*. Springer Nature, 2022.

Struchalin, G., Zagorovskii, Y. A., Kovlakov, E., Straupe, S., and Kulik, S. Experimental estimation of quantum state properties from classical shadows. *PRX Quantum*, 2(1):010307, 2021.

Takahashi, M. and Suzuki, M. One-dimensional anisotropic heisenberg model at finite temperatures. *Progress of theoretical physics*, 48(6):2187–2209, 1972.

Tang, P., Ramaiah, C., Wang, Y., Xu, R., and Xiong, C. Proposal learning for semi-supervised object detection. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 2291–2301, 2021.

Tang, Y., Xiong, H., Yang, N., Xiao, T., and Yan, J. Towards LLM4QPE: Unsupervised pretraining of quantum property estimation and a benchmark. In *The Twelfth International Conference on Learning Representations*, 2024.

Tarvainen, A. and Valpola, H. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. *Advances in neural information processing systems*, 30, 2017.

Van Engelen, J. E. and Hoos, H. H. A survey on semi-supervised learning. *Machine learning*, 109(2):373–440, 2020.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

Verma, V., Kawaguchi, K., Lamb, A., Kannala, J., Solin, A., Bengio, Y., and Lopez-Paz, D. Interpolation consistency training for semi-supervised learning. *Neural Networks*, 145:90–106, 2022.

Vidal, G. Efficient classical simulation of slightly entangled quantum computations. *Physical review letters*, 91(14): 147902, 2003.

Wang, H., Weber, M., Izaac, J., and Lin, C. Y.-Y. Predicting properties of quantum systems with conditional generative models. *arXiv preprint arXiv:2211.16943*, 2022.

Wang, X., Kihara, D., Luo, J., and Qi, G.-J. Enaet: A self-trained framework for semi-supervised and supervised learning with ensemble transformations. *IEEE Transactions on Image Processing*, 30:1639–1647, 2020.

Wu, D., Wang, L., and Zhang, P. Solving statistical mechanics using variational autoregressive networks. *Physical review letters*, 122(8):080602, 2019.

Wu, H. and Prasad, S. Semi-supervised deep learning using pseudo labels for hyperspectral image classification. *IEEE Transactions on Image Processing*, 27(3):1259–1270, 2017.

Wu, Y.-D., Zhu, Y., Bai, G., Wang, Y., and Chiribella, G. Quantum similarity testing with convolutional neural networks. *Physical Review Letters*, 130(21):210601, 2023a.

Wu, Y.-D., Zhu, Y., Wang, Y., and Chiribella, G. Learning and discovering quantum properties with multi-task neural networks. *arXiv preprint arXiv:2310.11807*, 2023b.

Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., and Philip, S. Y. A comprehensive survey on graph neural networks. *IEEE transactions on neural networks and learning systems*, 32(1):4–24, 2020.

Xiao, T., Huang, J., Li, H., Fan, J., and Zeng, G. Intelligent certification for quantum simulators via machine learning. *npj Quantum Information*, 8(1):138, 2022.

Xie, Q., Dai, Z., Hovy, E., Luong, T., and Le, Q. Unsupervised data augmentation for consistency training. *Advances in neural information processing systems*, 33: 6256–6268, 2020a.

Xie, Q., Luong, M.-T., Hovy, E., and Le, Q. V. Self-training with noisy student improves imagenet classification. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 10687–10698, 2020b.

Xu, M., Zhang, Z., Hu, H., Wang, J., Wang, L., Wei, F., Bai, X., and Liu, Z. End-to-end semi-supervised object detection with soft teacher. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 3060–3069, 2021.

Yi, X., Walia, E., and Babyn, P. Generative adversarial network in medical imaging: A review. *Medical image analysis*, 58:101552, 2019.

Zhai, X., Oliver, A., Kolesnikov, A., and Beyer, L. S4l: Self-supervised semi-supervised learning. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 1476–1485, 2019.

Zhang, B., Wang, Y., Hou, W., Wu, H., Wang, J., Okumura, M., and Shinozaki, T. Flexmatch: Boosting semi-supervised learning with curriculum pseudo labeling. *Advances in Neural Information Processing Systems*, 34: 18408–18419, 2021a.

Zhang, J. and Sarovar, M. Identification of open quantum systems from observable time traces. *Physical Review A*, 91(5):052121, 2015.

Zhang, L., Chen, Z., and Fei, S.-M. Entanglement verification with deep semisupervised machine learning. *Physical Review A*, 108(2):022427, 2023.

Zhang, T., Sun, J., Fang, X.-X., Zhang, X.-M., Yuan, X., and Lu, H. Experimental quantum state measurement with classical shadows. *Physical Review Letters*, 127(20): 200501, 2021b.

Zhang, Y.-H. and Di Ventra, M. Transformer quantum state: A multipurpose model for quantum many-body problems. *Physical Review B*, 107(7):075147, 2023.

Zhu, Y., Wu, Y.-D., Bai, G., Wang, D.-S., Wang, Y., and Chiribella, G. Flexible learning of quantum states with generative query neural networks. *Nature Communications*, 13(1):6222, 2022.

A. Related Work

A.1. Learning Classical Representations of Quantum States

Tensor networks (Shi et al., 2006) employ tensor decomposition to represent the quantum state as a series of small tensors. These method attempt to approximately store the low-entangled quantum states and simulates their evolution classically (Orús, 2019). The matrix product state (MPS) (Orús, 2014) is responsible for representing the one-dimensional array of qubits, whereas the projected entangled pair state (PEPS) (Corboz, 2016) generalizes the MPS from a one-dimensional array to an arbitrary graph. Recently, the classical shadow protocol (Huang et al., 2020) suggests to use random measurements to characterize the quantum properties, facilitating applications in state tomography (Hu et al., 2023), direct fidelity estimation (Struchalin et al., 2021), state function prediction (Zhang et al., 2021b).

Learning-based methods have recently gained attention, aiming to use neural networks as universal function approximators for various nonlinear properties of quantum systems. A multitude of models has emerged, including RBM-based (Carleo & Troyer, 2017), RNN-based (Carrasquilla et al., 2019), CNN-based (Wu et al., 2019; Sharir et al., 2020; Wu et al., 2023a), and transformer-based (Cha et al., 2021; Zhang & Di Ventra, 2023; Xiao et al., 2022) approaches. The specific choice of neural network depends on the topological structure of the corresponding physical problem. These methods are generally empirical and experimental.

Notably, recent advanced literature has started incorporating physical inductive biases into learning. For instance, Chen et al. (2023) integrates neural networks with tensor networks, where tensor networks preserve the physics structure and sign structure in the learned representations of quantum systems. Another study (Tang et al., 2024) introduces a pre-training paradigm that allows the model to effectively approximate the wave function of quantum states during the pre-training phase, with the learned parameters being highly effective for predicting other quantum properties. The authors of Qian et al. (2023) suggest to encode the circuit structure of the studied quantum system as input, significantly enhancing the model's ability to differentiate between various quantum states. Our method can serve as a universal paradigm for semi-supervised learning for quantum data. All these learning-based methods can be adapted within our framework to better suit the more commonly encountered scenario of sparsely labeled quantum data available.

A.2. Semi-Supervised Leaning in NLP and CV

In the realms of Natural Language Processing (NLP) and Computer Vision (CV), semi-supervised learning has emerged as a pivotal technique due to its ability to leverage both labeled and unlabeled data (Chen et al., 2022; Søgaard, 2022). This approach is particularly beneficial in scenarios where acquiring comprehensive labeled datasets is costly or unfeasible (Van Engelen & Hoos, 2020).

In NLP, semi-supervised methods have significantly advanced tasks such as language modeling (Peters et al., 2017) and sentiment analysis (Khan et al., 2017). One key strategy is self-training (Zhai et al., 2019), where models initially trained on labeled data are used to annotate and subsequently retrain on unlabeled data. This iterative cycle amplifies the training dataset, enhancing the model's learning capability. For example, the ULMFiT model (Howard & Ruder, 2018) has successfully adopted this approach, demonstrating remarkable efficacy in text classification. Additionally, transformer-based models like BERT (Devlin et al., 2019) and GPT (Brown et al., 2020), pre-trained on extensive unlabeled corpora, have set new benchmarks in various NLP tasks by fine-tuning on smaller, task-specific labeled datasets.

In CV, semi-supervised learning has revolutionized fields such as image classification (Wu & Prasad, 2017; Liu et al., 2020a) and object detection (Tang et al., 2021; Xu et al., 2021). Techniques like pseudo-labeling (Zhang et al., 2021a), where labels generated by a model are used to augment training data, have been widely implemented. Models such as Mean Teacher (Tarvainen & Valpola, 2017) and Noisy Student (Xie et al., 2020b) exemplify this approach, achieving high accuracy in image classification tasks. Furthermore, Generative Adversarial Networks (GANs) (Goodfellow et al., 2014) have been leveraged for data augmentation in domains like medical imaging (Yi et al., 2019) for addressing the scarcity of labeled data.

B. More Basics of Quantum Computing

In the main text of the paper, we only introduce the mathematical formulation of Pauli-X operator and its eigen-decomposition. Below we provide more details about the other two Pauli operators. For a comprehensive discussion, we refer the readers who are interested in quantum computing and quantum information to the Section 2.1 of the book (Nielsen & Chuang, 2010).

A single qubit is a vector $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ parameterized by two complex numbers satisfying $|\alpha|^2 + |\beta|^2 = 1$. Operations on a qubit must preserve this norm, and thus are described by 2×2 unitary matrices. Of these, some of the most important are the Pauli operators; it is useful to list them again here:

$$\boldsymbol{\sigma}^x \equiv \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \boldsymbol{\sigma}^y \equiv \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \quad \boldsymbol{\sigma}^z \equiv \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}. \tag{7}$$

One could do some linear algebras and check that $|0\rangle = \left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right]$ and $|1\rangle = \left[\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right]$ are the eigenvectors of $\boldsymbol{\sigma}^z$, $|+\rangle = \frac{1}{\sqrt{2}}\left[\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right]$ and $|-\rangle = \frac{1}{\sqrt{2}}\left[\begin{smallmatrix} 1 \\ -1 \end{smallmatrix}\right]$ are the eigenvectors of $\boldsymbol{\sigma}^x$, $|\boldsymbol{i}_+\rangle = \frac{1}{\sqrt{2}}\left[\begin{smallmatrix} 1 \\ \boldsymbol{i} \end{smallmatrix}\right]$ and $|\boldsymbol{i}_-\rangle = \frac{1}{\sqrt{2}}\left[\begin{smallmatrix} 1 \\ -\boldsymbol{i} \end{smallmatrix}\right]$ are the eigenvectors of $\boldsymbol{\sigma}^y$. The same qubit can be decomposed in to different orthonormal basis. For example,

$$\begin{aligned} |\psi\rangle &= \alpha|0\rangle + \beta|1\rangle \\ &= \frac{1}{\sqrt{2}}(\alpha+\beta)|+\rangle + \frac{1}{\sqrt{2}}(\alpha-\beta)|-\rangle \\ &= \frac{1}{\sqrt{2}}(\alpha-\beta\boldsymbol{i})|\boldsymbol{i}_+\rangle + \frac{1}{\sqrt{2}}(\alpha+\beta\boldsymbol{i})|\boldsymbol{i}_-\rangle. \end{aligned} \tag{8}$$

The density matrix representations of these eigen-states fulfill the criteria for projective measurements. It means that $\{|0\rangle\langle 0|, |1\rangle\langle 1|\}$, $\{|+\rangle\langle +|, |-\rangle\langle -|\}$ and $\{|\boldsymbol{i}_+\rangle\langle \boldsymbol{i}_+|, |\boldsymbol{i}_-\rangle\langle \boldsymbol{i}_-|\}$ are all the projective measurements. We employ these as random Pauli measurements to observe the quantum system and gather the resultant measurement data.

We may consider a system of L qubits. It can be described by the so called *wave function*:

$$|\boldsymbol{\Phi}\rangle = \sum_{\sigma_1=1}^{M} \cdots \sum_{\sigma_L=1}^{M} \boldsymbol{\Psi}(\sigma_1, \ldots, \sigma_L)|\sigma_1, \ldots, \sigma_L\rangle, \tag{9}$$

where $\boldsymbol{\Psi} : \mathbb{Z}^L \to \mathbb{C}$ maps a fixed configuration $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_L)$ of L qubits to a complex number which is the amplitude satisfying $\sum_{\sigma_1=1}^{K} \cdots \sum_{\sigma_L=1}^{K} |\boldsymbol{\Psi}(\sigma_1, \ldots, \sigma_L)|^2 = 1$, and $\sigma_i \in \{1, \ldots, K\}$ is one of the K possible outcomes by performing quantum measurement on the i-th qubit. It is formulated in a complex Hilbert space where the vector representation of the quantum state $|\boldsymbol{\Phi}\rangle \in \mathbb{C}^{K^L}$ and its density matrix $|\boldsymbol{\Phi}\rangle\langle\boldsymbol{\Phi}| \in \mathbb{C}^{K^L \times K^L}$, which becomes astronomical for large L.

C. More Details About the Dataset Construction

C.1. Heisenberg Model

In our study, we perform simulated experiments on the one-dimensional transverse-field Ising model, scaling up to 50 qubits. Direct simulation of such a large-qubit quantum system, involving manipulation and storage of maximal 2^{50}-dimensional complex vectors, is impractical due to computational constraints. To overcome this, we utilize Matrix Product State (MPS) methods. Specifically, we leverage the Julia implementation provided by ITensor Fishman et al. (2022), for representing the quantum state and simulating its temporal evolution.

In order to computing the time evolution of a quantum state under the dynamics of a Hamiltonian H, we adopt the technique named as "time evolving block decimation" (TEBD) (Vidal, 2003). The idea is to decompose the time-evolution operator into a circuit of quantum gates (two-dimensional unitaries) using the Trotter-Suzuki approximation (Hatano & Suzuki, 2005) and to apply these gates in a controlled way to an MPS. Consider the Hamiltonian we use to generate the dataset given in Eq. 4, and suppose that the total evolution time T is discreted as N_T small time steps $\tau = T/N_T$, the Trotter decomposition is given as

$$\begin{aligned} e^{-\boldsymbol{i}\tau H} \approx & e^{-\boldsymbol{i}\boldsymbol{\sigma}_1^z\boldsymbol{\sigma}_2^z J\tau/2} e^{-\boldsymbol{i}\boldsymbol{\sigma}_2^z\boldsymbol{\sigma}_3^z J\tau/2} \cdots e^{-\boldsymbol{i}\boldsymbol{\sigma}_{L-1}^z\boldsymbol{\sigma}_L^z J\tau/2} e^{-\boldsymbol{i}\boldsymbol{\sigma}_1^x Jg\tau/2} \cdots e^{-\boldsymbol{i}\boldsymbol{\sigma}_L^x Jg\tau/2} \times \\ & e^{-\boldsymbol{i}\boldsymbol{\sigma}_L^x Jg\tau/2} \cdots e^{-\boldsymbol{i}\boldsymbol{\sigma}_1^x Jg\tau/2} e^{-\boldsymbol{i}\boldsymbol{\sigma}_{L-1}^z\boldsymbol{\sigma}_L^z J\tau/2} \cdots e^{-\boldsymbol{i}\boldsymbol{\sigma}_1^z\boldsymbol{\sigma}_2^z J\tau/2} + O\left(\tau^3\right). \end{aligned} \tag{10}$$

The error in the above decomposition is of order τ^3, so that will be the error accumulated per time step. Because of the time-step error, one takes τ to be small and then applies the above set of operators to an MPS as a single sweep, then does N_T sweeps to evolve for a total time T. The total error will therefore scale as τ^2 with this scheme, though other sources of error may dominate for long times, or very small τ such as truncation errors. In our simuation, the initial state are set to be the a product state alternating up ($|0\rangle$) and down ($|1\rangle$). The truncation error threshold is fixed as $1 \times e^{-12}$ to enable a good approximation of the quantum state. Note that we only conduct unitary evolution for MPS time evolution since the non-Hermitian evolution is relatively hard to be implemented based on linear tensor networks.

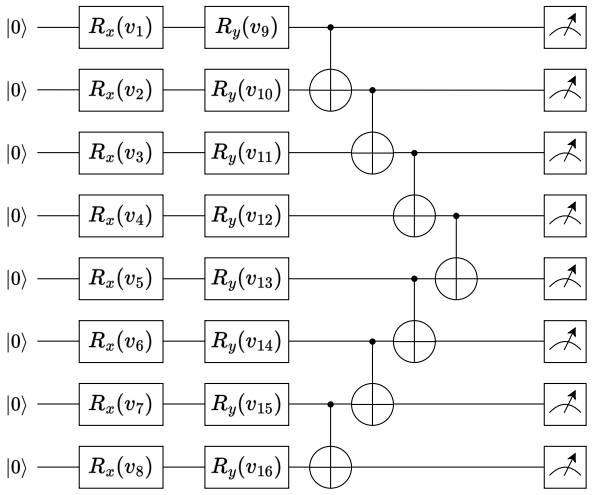

Figure 8. The quantum circuit used for mapping the classical images into the quantum states.

C.2. VQC Model

In Sec. 4.2, we delve into the methodology of encoding images of classes 0 and 1 into quantum states using a Variational Quantum Circuit (VQC). To elucidate this process, we present the layout of the VQC as utilized for dataset generation in Fig. 8. This circuit, which comprises a total of 8 qubits, is conceptually akin to the embedding layer in the Quantum Convolutional Neural Network (QCNN) as implemented by Hur et al. (2022). Notably, an entanglement layer, densely populated with CNOT gates, is appended to this configuration. Specifically, followed by Hur et al. (2022), each 28×28 grayscale image is first transformed into a 16-dimensional feature vector $\boldsymbol{v}$ using Principal Component Analysis (PCA). Within this vector, the initial 8 elements are designated as parameters for the R_x gates, while the subsequent 8 elements are allocated as parameters for the R_y gates, where

$$R_x(\theta) = e^{-i\theta\boldsymbol{\sigma}_x/2} = \begin{bmatrix} \cos\frac{\theta}{2} & -i\sin\frac{\theta}{2} \\ -i\sin\frac{\theta}{2} & \cos\frac{\theta}{2} \end{bmatrix}, \qquad R_y(\theta) = e^{-i\theta\boldsymbol{\sigma}_y/2} = \begin{bmatrix} \cos\frac{\theta}{2} & -\sin\frac{\theta}{2} \\ \sin\frac{\theta}{2} & \cos\frac{\theta}{2} \end{bmatrix}. \tag{11}$$

Then we perform a series of CNOT gates to generate an entangled quantum state, with the two-qubit CNOT gate written as

$$\text{CNOT} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}. \tag{12}$$

Ultimately, we execute quantum measurements and gather the resulting measurement records to assemble our dataset. The simulation of the Variational Quantum Circuit (VQC) and the subsequent quantum measurements are conducted using the Pennylane toolkit (Bergholm et al., 2018).

D. Detailed Experimental Settings

D.1. Baselines

Kernel Methods (Huang et al., 2022). We utilize Gaussian Kernel and Neural Tangent Kernel (NTK) to transform shadow embeddings into a feature vector with dimensions $M \times 8L$ for all samples in both training and testing datasets. A grid search is performed to identify the optimal regularization strength, with candidate values uniformly distributed on a logarithmic scale from 0.001 to 100. We employ a 5-fold cross-validation strategy on the training dataset and present the predictive performance of the model that achieves the highest accuracy on the test dataset.

RNN (Carrasquilla et al., 2019). We adopt a Gated Recurrent Unit (GRU) (Cho et al., 2014) to project the shadow embeddings into a higher-dimensional feature space. A subsequent Feed-Forward Neural Network (FFN) is appended to the final layer to compress the feature dimensions to match the number of target classes. Benefit from the inherent

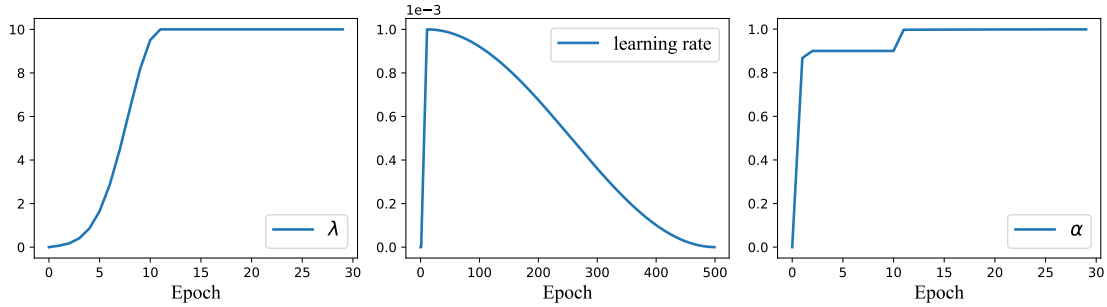

Figure 9. The pre-defined schedule of the consistency weight λ (left), the learning rate (middle) and the EMA decay coefficient α (right).

auto-regressive properties, RNN has been widely adopted to represent the wave functions and predict the properties of quantum states (Carrasquilla, 2020). The hidden dimension of GRU is set to be 128.

LLM4QPE (Tang et al., 2024). LLM4QPE presents a versatile approach for predicting various properties of quantum systems through a pre-training procedure that maximizes a likelihood function based on discrete measurement outcomes. For a fair comparison, we bypass the pre-training phase and configure the model with 4 heads, 2 layers, and a hidden dimension of 128.

CNN and QCNN (Hur et al., 2022). Following the pre-processing protocol in Hur et al. (2022), we apply PCA to condense the 28×28 images into a 16-dimensional feature vector prior to model input. For the CNN, we set the convolutional kernel size to 2, stride to 1, and padding to 1. The number of free parameters used is 34. The QCNN model is equipped with 8 qubits and utilizes a hardware-efficient ansatz, comprising 26 trainable parameterized gates with convolutional circuit 2 in Hur et al. (2022).

D.2. Our Proposed SSL4Q and SSL4Q sup.

The primary distinction between our SSL4Q model and existing supervised baselines lies in our approach to harnessing the potential of unlabeled quantum data. We achieve this through the minimization of consistency loss during model training. The key to this process are two hyper-parameters: the standard deviation (std) of Gaussian noise and the consistency weight λ. Based on empirical evidence, we set the Gaussian noise mean to 0 and std to 0.25, which consistently yields optimal predictive accuracy across various scenarios. The training duration is established at 500 epochs.

We employ a tailored schedule for the consistency weight λ, learning rate, and the Exponential Moving Average (EMA) decay coefficient α, as depicted in the Fig. 9. Initially, λ is increased slowly to mitigate the impact of insufficient supervision early in training, preventing significant training loss fluctuations or convergence issues. λ experiences an exponential rise in the first 10 epochs, peaking at a value of 10. The learning rate ascends to a maximum of 0.001 and subsequently follows a cosine-like decline. The α coefficient adjusts in three phases over the training process. The Adam optimizer is used for optimization.

For SSL4Q, the architecture comprises 4 heads, 2 layers, and a hidden dimension of 128. The supervised variant of SSL4Q shares the same structure without data augmentation and the parameters are optimized by minimizing supervised loss.

E. Additional Numerical Results

E.1. Impact of Different Quantum Data Embedding Strategies

Our proposed shadow embedding includes both the measurement string and the corresponding measurement operator, as discussed in Sec. 3.3 and illustrated in Fig. 2 and 3. Firstly, we consider the scenario where the measurement operator information is removed from the shadow embedding (denoted as Shadow Embedding w/o Meas. Op.). This is akin to the embedding strategies discussed in [1,2], where the model's input is the measurement string, with each element being discrete (ranging from 0 to $K-1$, where K is the number of possible outcomes when measuring a single qubit, where $K=6$ in our study). We conducted experiments on the Heisenberg model with $M=128$ and $\eta=0.6$ for $L=20$. The results are averaged over 3 runs with different random seed initializations and are listed in Fig. 10. For comparison, we also list the

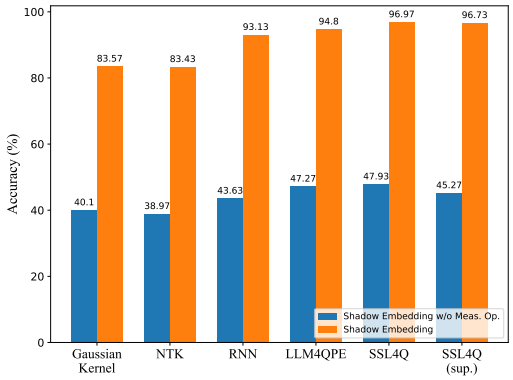

Figure 10. Comparison of different quantum data embedding strategies with and without measurement operator information.

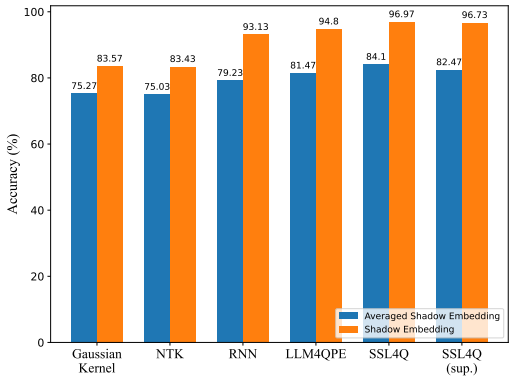

Figure 11. Comparison of different quantum data embedding strategies with and without average operation.

Table 4. Classification accuracy in predicting different classes of Heisenberg model quantum system, with $\eta = 0.9$ and $M = 32$. In the table, 'w/o shadow embedding' denotes we substitute the shadow embedding with discrete measurement outcomes, while 'w/o PEA' represents we replace the PEA Layer with an LSTM layer. The best results are highlighted in **bold**.

Method	$L = 20$	$L = 30$	$L = 40$	$L = 50$
SSL4Q (w/o shadow embedding)	52.57	54.60	53.23	51.37
SSL4Q (w/o PEA)	73.30	71.47	70.80	71.23
SSL4Q	**84.53**	**80.13**	**78.03**	**73.40**

performance of the original shadow embedding accordingly. The experimental results demonstrate that relying solely on the discrete measurement string as input significantly decreases the prediction performance.

Furthermore, we study averaging the shadow embedding $X \in \mathbb{R}^{M\times L}$ (M is the number of measurements and L is the system size) along the first axis before inputting it into the model, transforming it into a vector $X' \in \mathbb{R}^{L}$ (denoted as Averaged Shadow Embedding). The results are reported in Fig. 11. It is evident that there is a decrease in evaluation performance on both the baselines and SSL4Q.

E.2. Other Results

We present additional ablation study for the model with and without shadow embedding and PEA layer and the results are listed in Tab. 4. Further investigations about the impacts of measurement uncertainties are illustrated in Fig. 12. The sensitivity to the data augmentation and consistency weight are reported in Fig. 13.

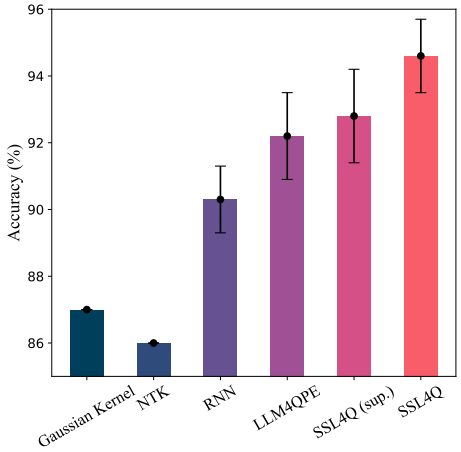

Figure 12. Classification accuracy in predicting different classes of Heisenberg model quantum system for a toy dataset generated by $t = 5$ms, $g = 1$, $L = 50$, $\eta = 0.9$, and $M = 16$, with varied $g \in \{0.5, 1.5\}$. The size of training dataset is equal to the testing dataset $N^{tr} = N^{te} = 100$. For the neural network based models, the accuracy is averaged over 5 runs with different random seeds.

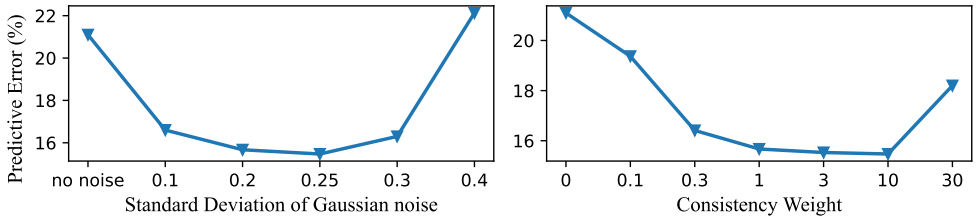

Figure 13. The SSL4Q's sensitivity w.r.t. the noise level and the (maximal) consistency weight. The predictive error is reported on the dataset with $L = 20$, $M = 32$ and $\eta = 0.9$.

F. Limitations

In this paper, we have confined our experiments within the typical semi-supervised learning framework of consistency regularization (Tarvainen & Valpola, 2017). More advanced methods, such as generative methods (Liu et al., 2020b) and pseudo-labeling methods (Wang et al., 2020), might also yield similar or improved predictive performance for quantum state classification with limited labeled data.

Our studies focus on classification tasks which subject to identify different types of quantum states. A complementary task is to estimate the properties of quantum states such as expectation values of observables, entanglement entropy of sub-systems and correlation functions. We do not consider the regression tasks of quantum properties estimation in this paper. This is mainly because limited labeled training samples may not fully represent the entire distribution of the whole data when the labels have continuous values. If the labeled samples do not cover the range of labels in the test set, it could lead to decreased model performance when predicting the properties of the unseen quantum data. The challenges of applying semi-supervised learning to regression tasks in quantum properties estimation could potentially be overcome by employing strategies such as data augmentation and interpolation (Verma et al., 2022), or active learning (Gao et al., 2020), but these are beyond the scope of this paper.

In future work, we aim to further explore these aspects. Additionally, future studies could attempt to devise specialized data augmentation techniques for quantum data to enhance model performance, as this has been proven to play a crucial role in semi-supervised learning (Xie et al., 2020a).