Neural Quantum Embedding:

Pushing the Limits of Quantum Supervised Learning

Quantum Techniques in Machine Learning | November 23rd 2023

Tak Hur (w/ Israel F. de Araujo & Daniel K. Park)

Yonsei University

Department of Statistics and Data Science

Quantum Supervised Learning

Given N labelled sample classical data,

$$\{(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)\} \overset{\text{iid}}{\sim} D$$

Find prediction function $f(x)$,

$$\min \mathbb{E}_{X,Y \sim D} |f(X) - Y|$$

Quantum Supervised Learning

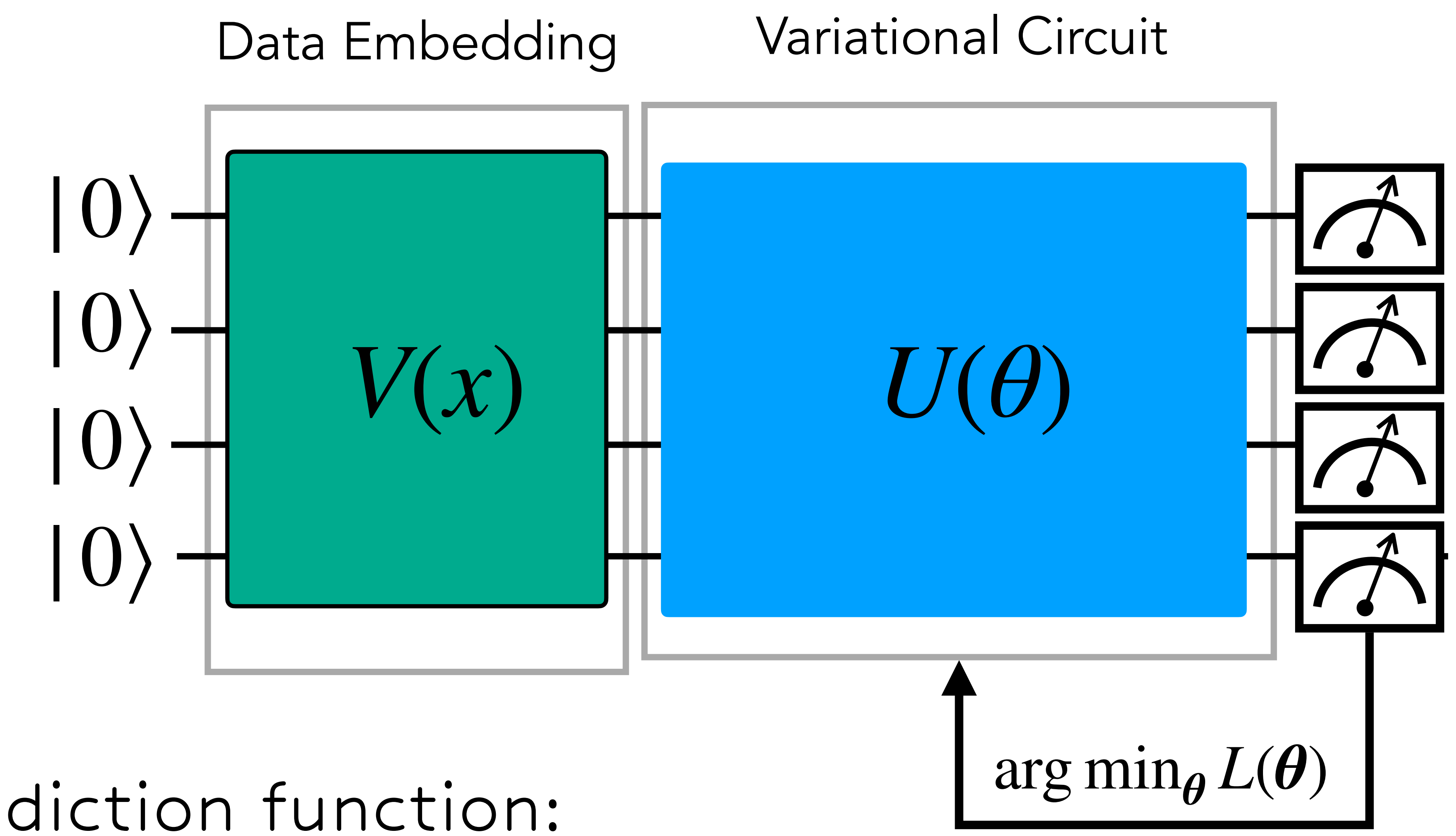
Given N labelled sample classical data,

$$\{(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)\} \overset{\text{iid}}{\sim} D$$

Find prediction function $f(x)$,

$$\min \mathbb{E}_{X,Y\sim D} |f(X) - Y|$$

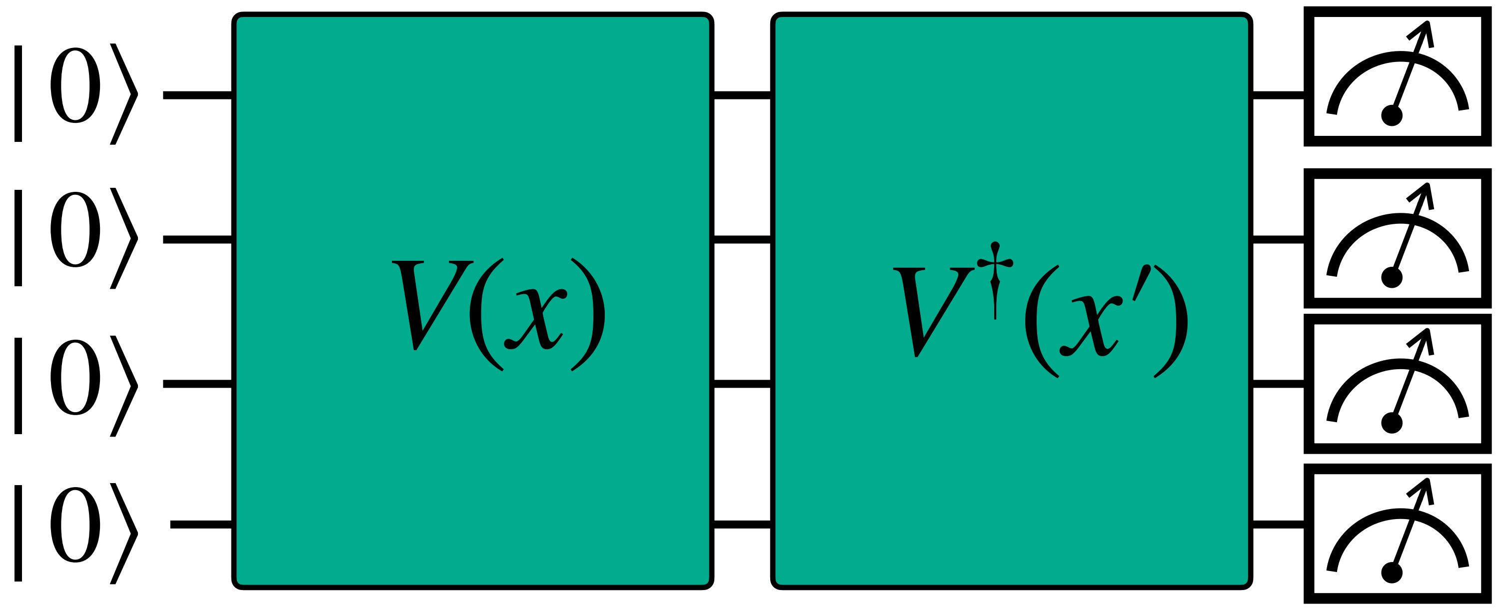
1. Quantum Neural Networks



Prediction function:

$$f(x;\theta) = \langle x | U(\theta)^\dagger O U(\theta) | x \rangle$$

2. Quantum Kernel Methods



Kernel function:

$$k(x, x') = |\langle x | x' \rangle|^2$$

*Note Quantum Embedding: $V(x)|0\rangle^{\otimes n} = |x\rangle$

Lower bound of Empirical Risk (QNN)

- Given data $\{(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)\} \overset{\text{iid}}{\sim} D$, Quantum Embedding $V(\cdot)$ determines the lower bound of Empirical Risk (Training error)

Lower bound of Empirical Risk (QNN)

- Given data $\{(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)\} \overset{\text{iid}}{\sim} D$, Quantum Embedding $V(\cdot)$ determines the lower bound of Empirical Risk (Training error)

Consider Binary Classification $y_i \in \{+1, -1\}$ & Linear Loss $l(f(x), y) = \frac{1}{2}|y - f(x)|$,

- Training QNN = Quantum State Discrimination with POVM $\{E_+(\theta), E_-(\theta)\}$, where $E_\pm(\theta) = \frac{1}{2}(I \pm U^\dagger(\theta) O U(\theta))$

Lower bound of Empirical Risk (QNN)

- Given data $\{(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)\} \overset{\text{iid}}{\sim} D$, Quantum Embedding $V(\cdot)$ determines the lower bound of Empirical Risk (Training error)

Consider Binary Classification $y_i \in \{+1, -1\}$ & Linear Loss $l(f(x), y) = \frac{1}{2}|y - f(x)|$,

- Training QNN = Quantum State Discrimination with POVM $\{E_+(\theta), E_-(\theta)\}$, where $E_\pm(\theta) = \frac{1}{2}(I \pm U^\dagger(\theta) O U(\theta))$

$$L_s(\theta) = \frac{1}{N}\left[\sum_{i=1}^{N^-} P(E_+(\theta) \,|\, x_i^-) + \sum_{i=1}^{N^+} P(E_-(\theta) \,|\, x_i^+)\right]$$

$$\geq \frac{1}{2} - D_{\mathrm{tr}}(p^-\rho^-, p^+\rho^+)$$

$$\rho^\pm = \frac{1}{N^\pm} \sum |x_i^\pm\rangle\langle x_i^\pm|$$

$$p^\pm = N^\pm / N$$

By having Large Trace Distance,

- Smaller Lower Bound of Empirical Risk (Training Error)
- Classification task becomes robust against noise

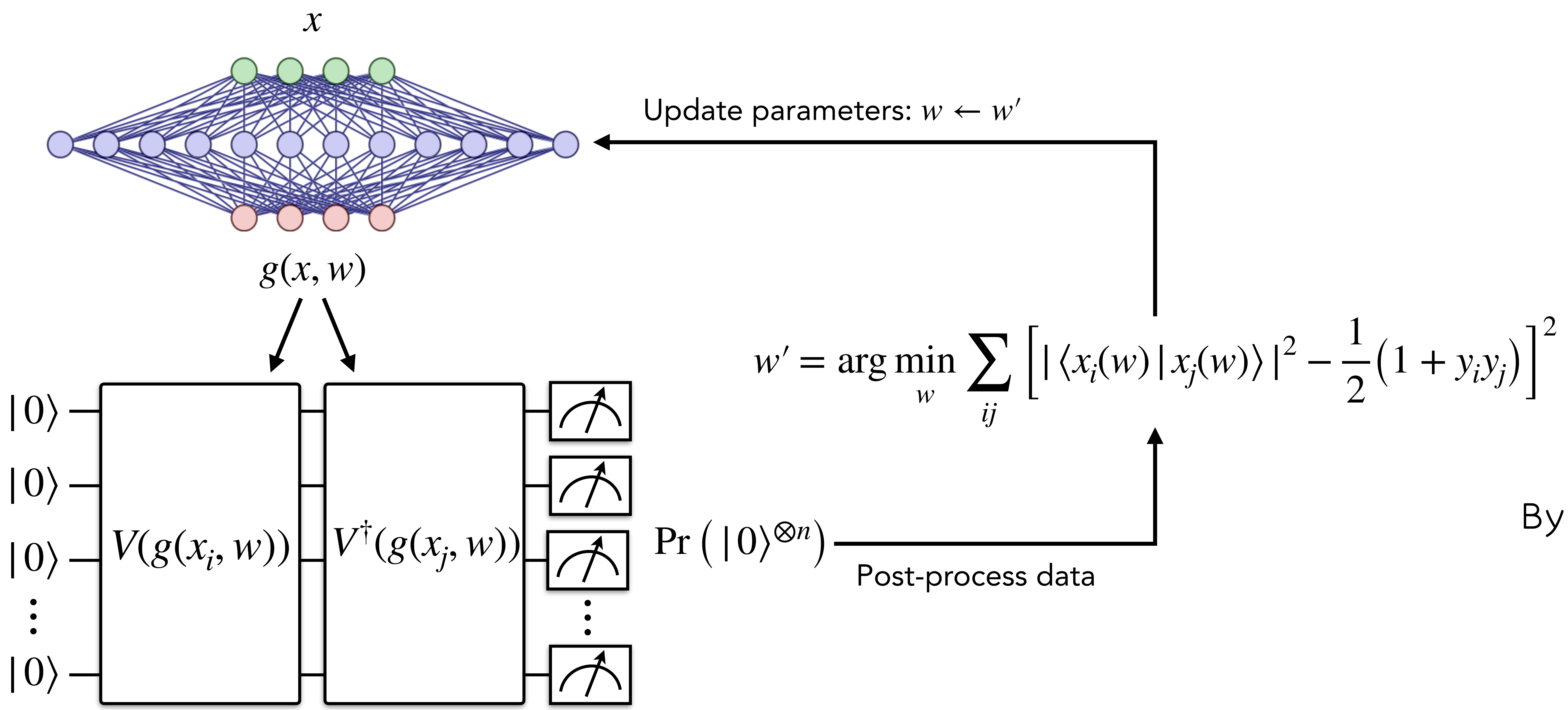
$$D_{\mathrm{tr}}(\Lambda(\rho_0), \Lambda(\rho_1)) \leq D_{\mathrm{tr}}(\rho_0, \rho_1)$$

By having Large Trace Distance,

- Smaller Lower Bound of Empirical Risk (Training Error)
- Classification task becomes robust against noise

$$D_{\mathrm{tr}}(\Lambda(\rho_0), \Lambda(\rho_1)) \leq D_{\mathrm{tr}}(\rho_0, \rho_1)$$

Neural Quantum Embedding (NQE)



By Choosing $m > m'$ for NQE

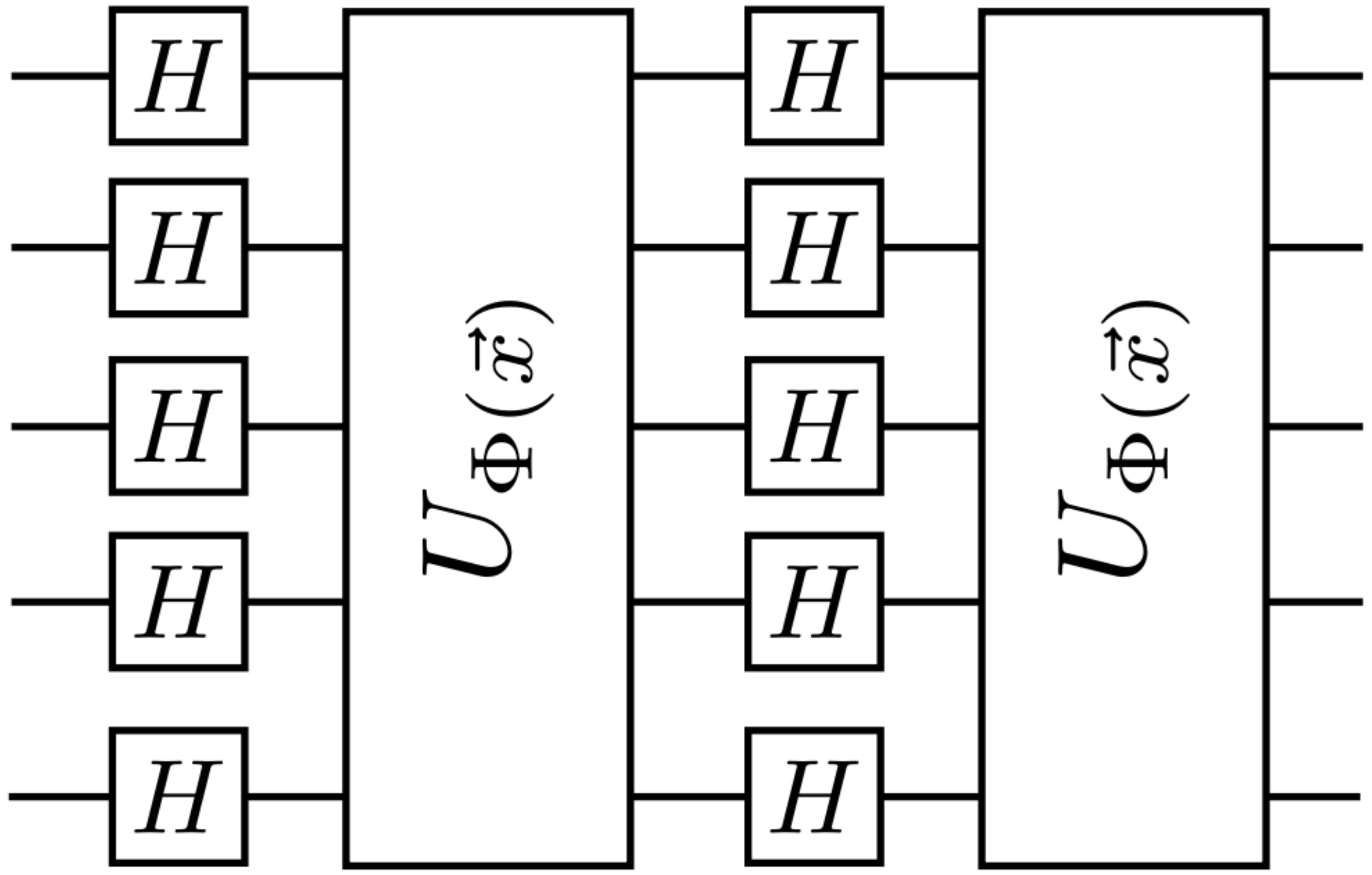
$g : \mathbb{R}^m \rightarrow \mathbb{R}^{m'}$

Experimental Results

Hamiltonian Encoding

- A popular example of the quantum feature map:

$\mathcal{U}(\boldsymbol{x}) =$



$$f(\boldsymbol{x}, \boldsymbol{\theta}) = \langle 0 | \left(\mathcal{U}^{\dagger}(\boldsymbol{x})\right)^{d} V^{\dagger}(\boldsymbol{\theta}) O V(\boldsymbol{\theta}) \left(\mathcal{U}(\boldsymbol{x})\right)^{d} | 0 \rangle$$

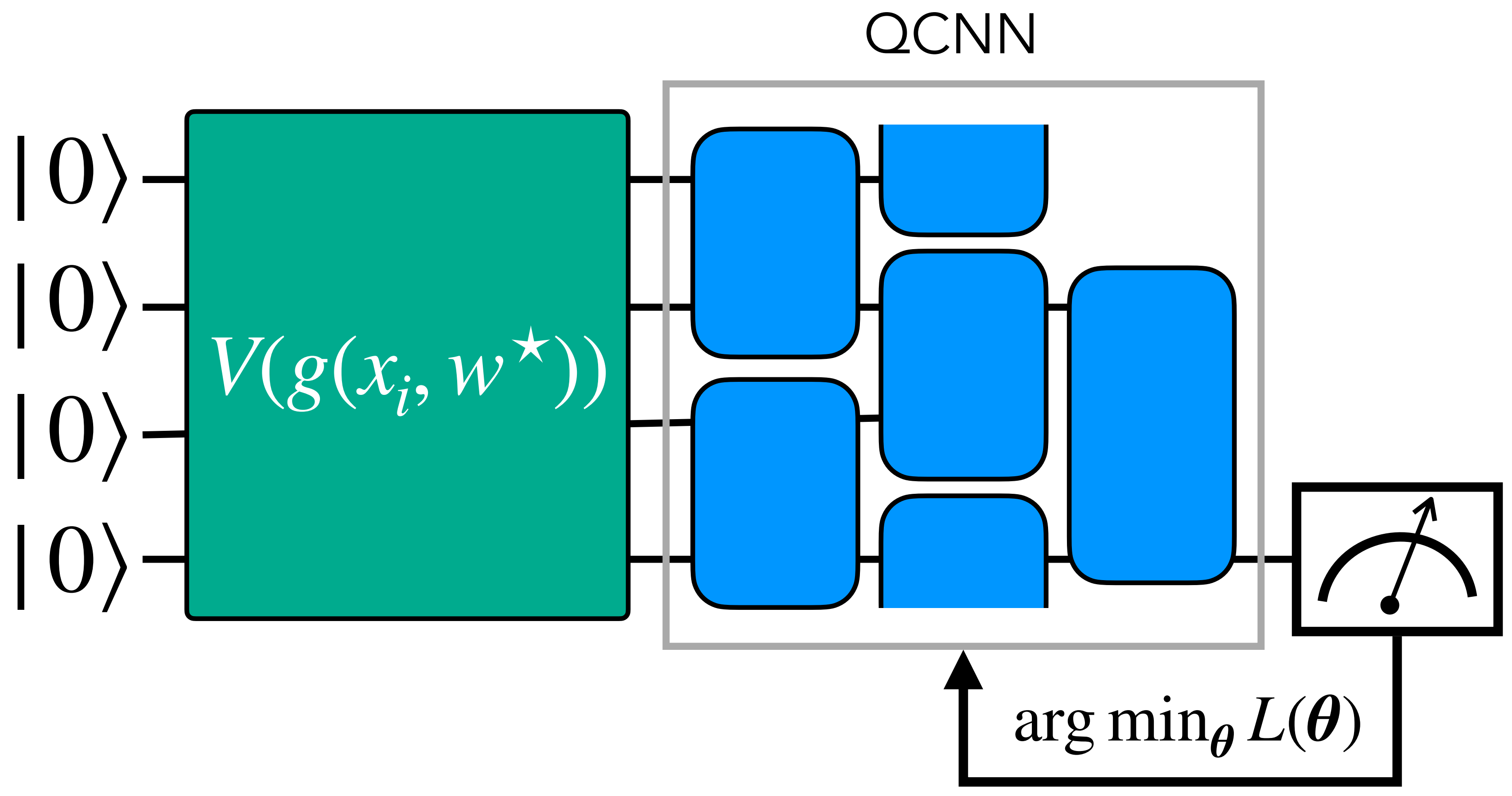
$$k(\boldsymbol{x}, \boldsymbol{y}) = | \langle 0 | \left(\mathcal{U}^{\dagger}(\boldsymbol{y})\right)^{d} \left(\mathcal{U}(\boldsymbol{x})\right)^{d} | 0 \rangle |^{2}$$

- $U_{\Phi}(\vec{x}) = \exp\left[i \left(\sum_{j} \phi_1(x_j) Z_j + \sum_{j<k} \phi_2(x_j, x_k) Z_j Z_k \right) \right]$ with some functions ϕ_1 and ϕ_2 (a.k.a ZZ feature map)
- Typical example: $\phi_1(x) = x$, $\phi_2(x, y) = (\pi - x)(\pi - y)$

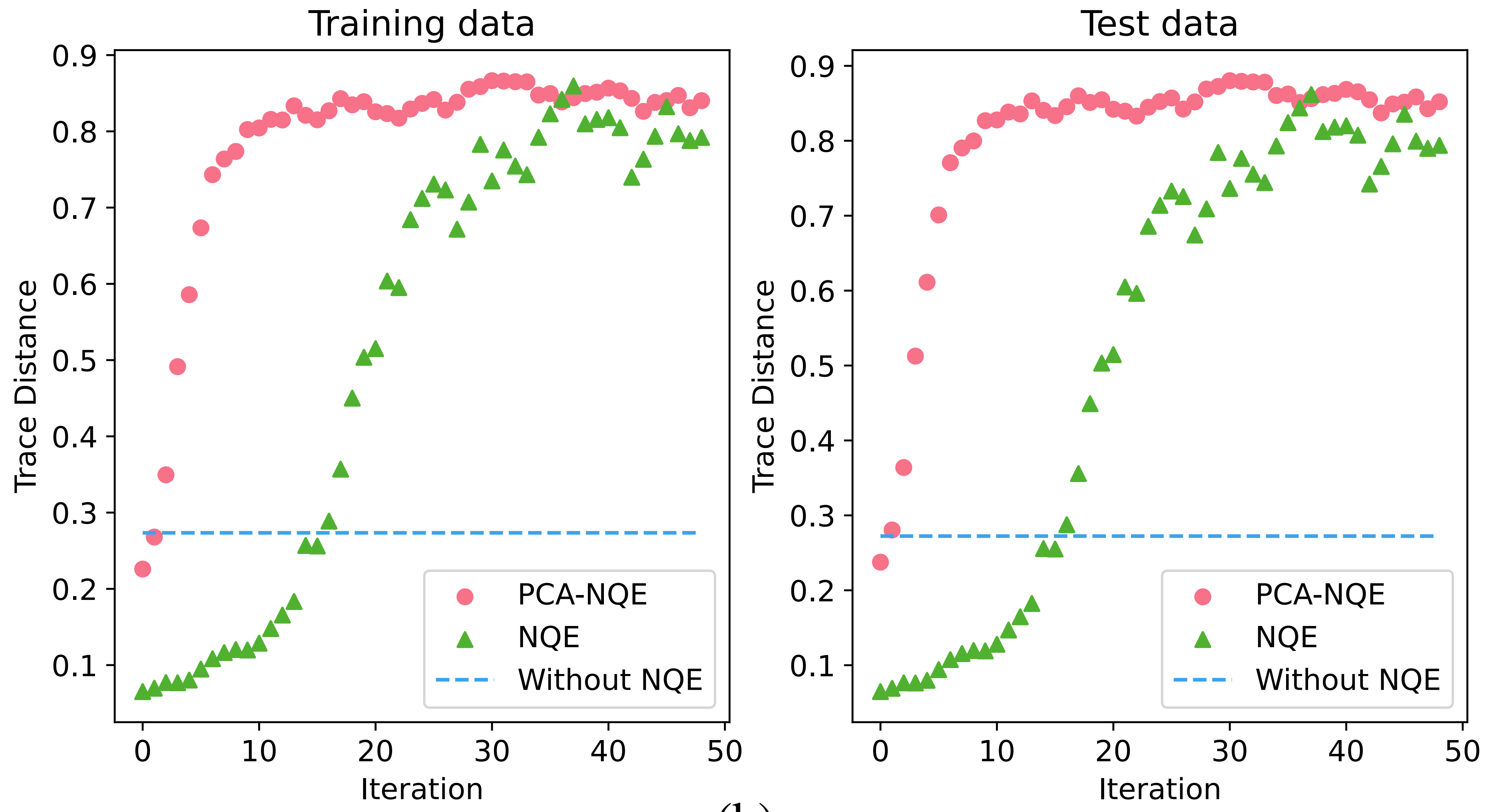
Havlíček et al. Nature 567, 209-212 (2019)
Abbas et al. Nature Comp. Sci. 1, 403-409 (2021)

Experimental Results

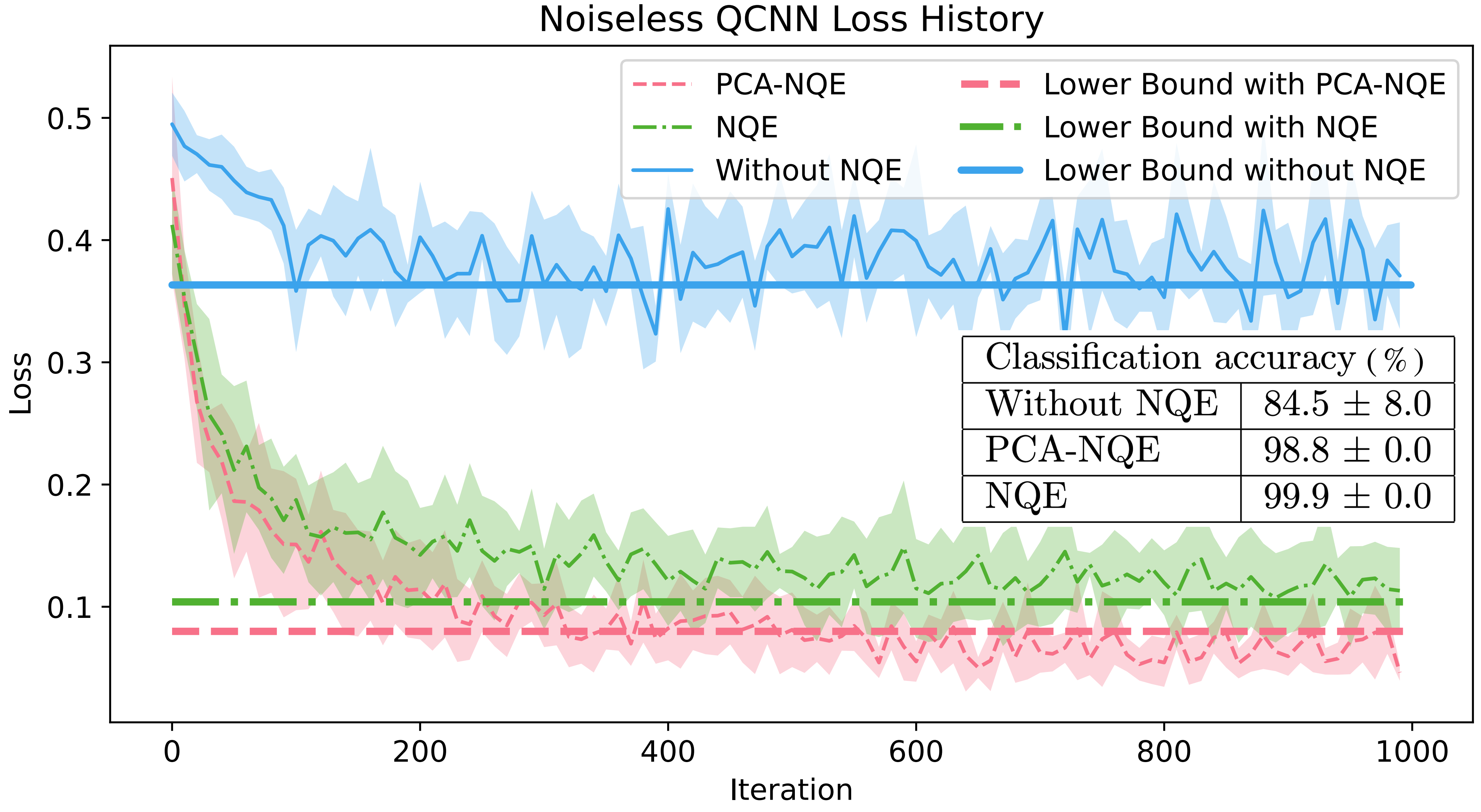
Training QCNN with and without NQE



(a)

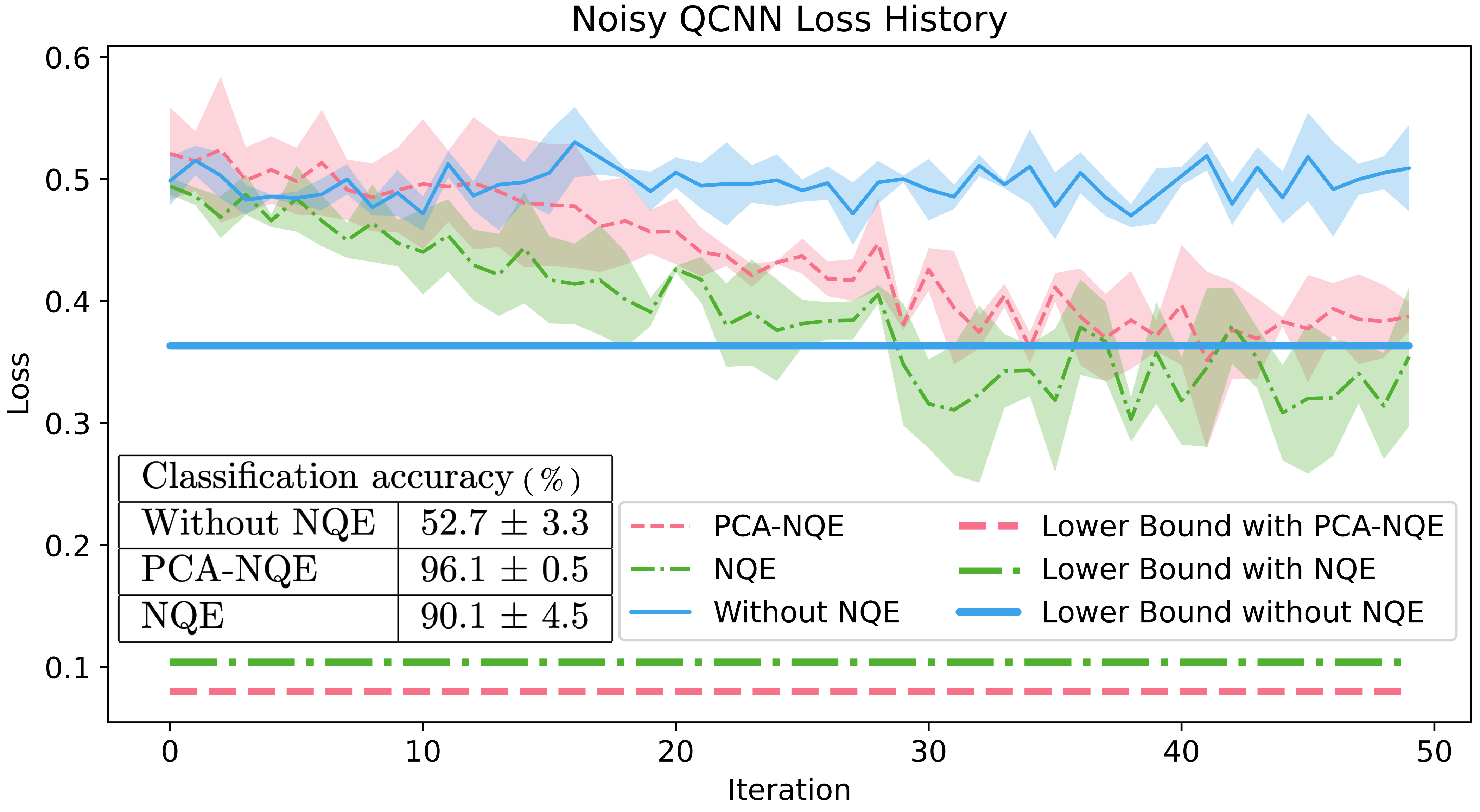


(b)



Classification accuracy (%)	
Without NQE	84.5 ± 8.0
PCA-NQE	98.8 ± 0.0
NQE	99.9 ± 0.0

(c)



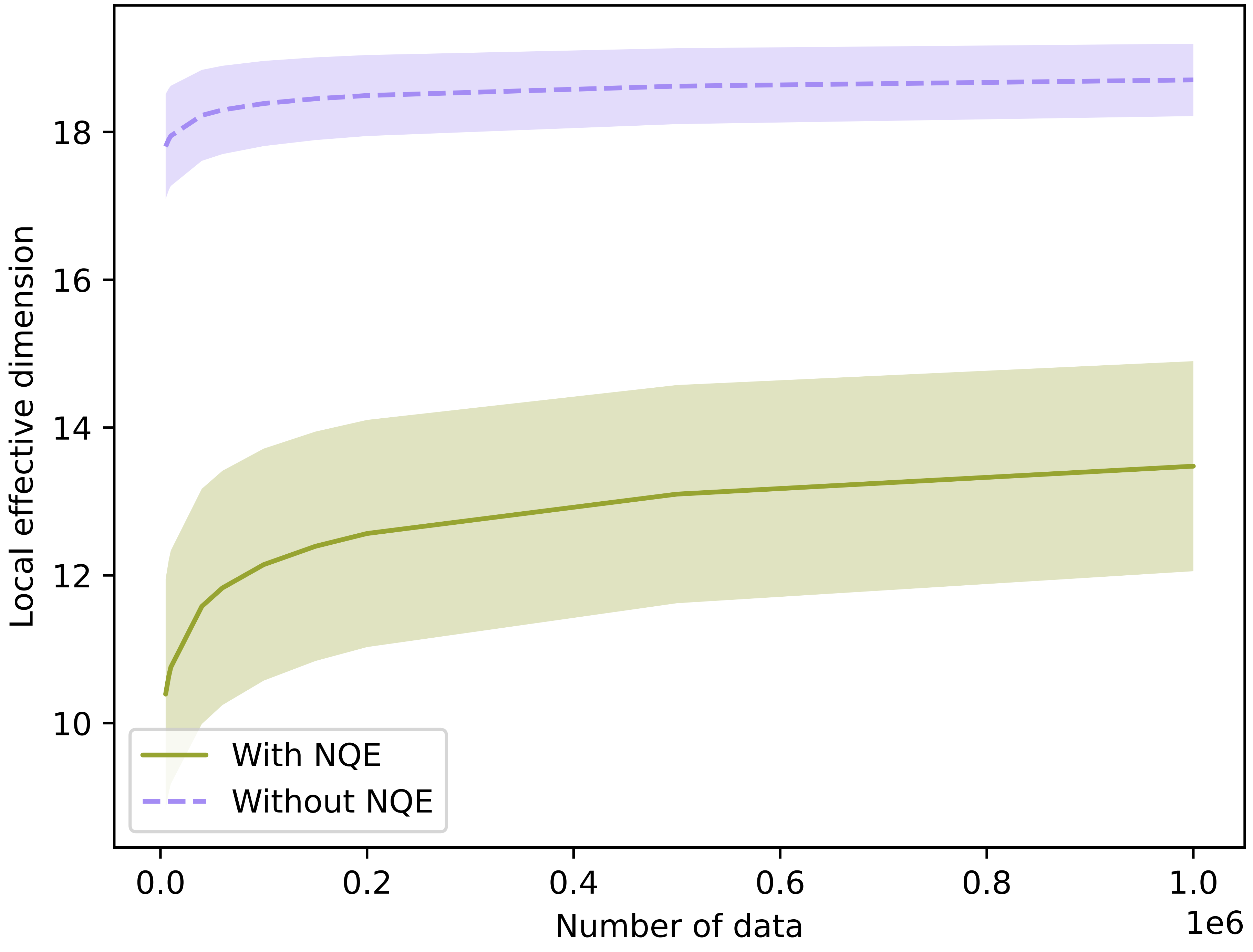
Classification accuracy (%)	
Without NQE	52.7 ± 3.3
PCA-NQE	96.1 ± 0.5
NQE	90.1 ± 4.5

(d)

Generalization Performance

1. Quantum Neural Networks

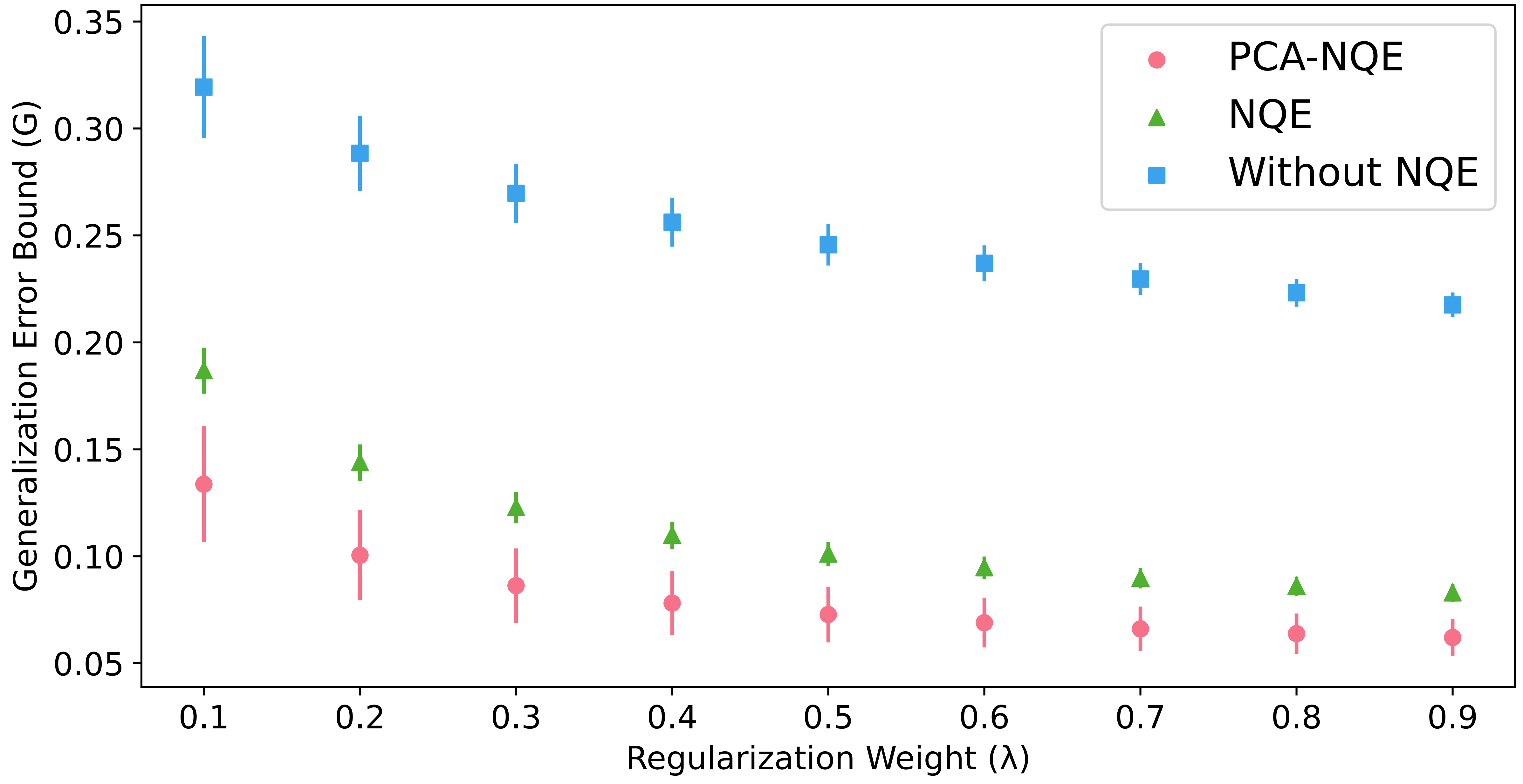
- Local Effective Dimension:
  Complexity metric for Learning Model
  (Abbas et al. arXiv:2112.04807)
- Positive Correlation with Generalization Error



2. Quantum Kernel Methods

$$\left| R(W) - R_N(W) \right| \le \mathcal{O}\left(\frac{||W||_F}{\sqrt{N}} \right)$$

where, $$W^* = \sum_{i=1}^{N} \sum_{j=1}^{N} y_i (K^Q + \lambda I)^{-1}_{i,j} |x_j\rangle\langle x_j|$$

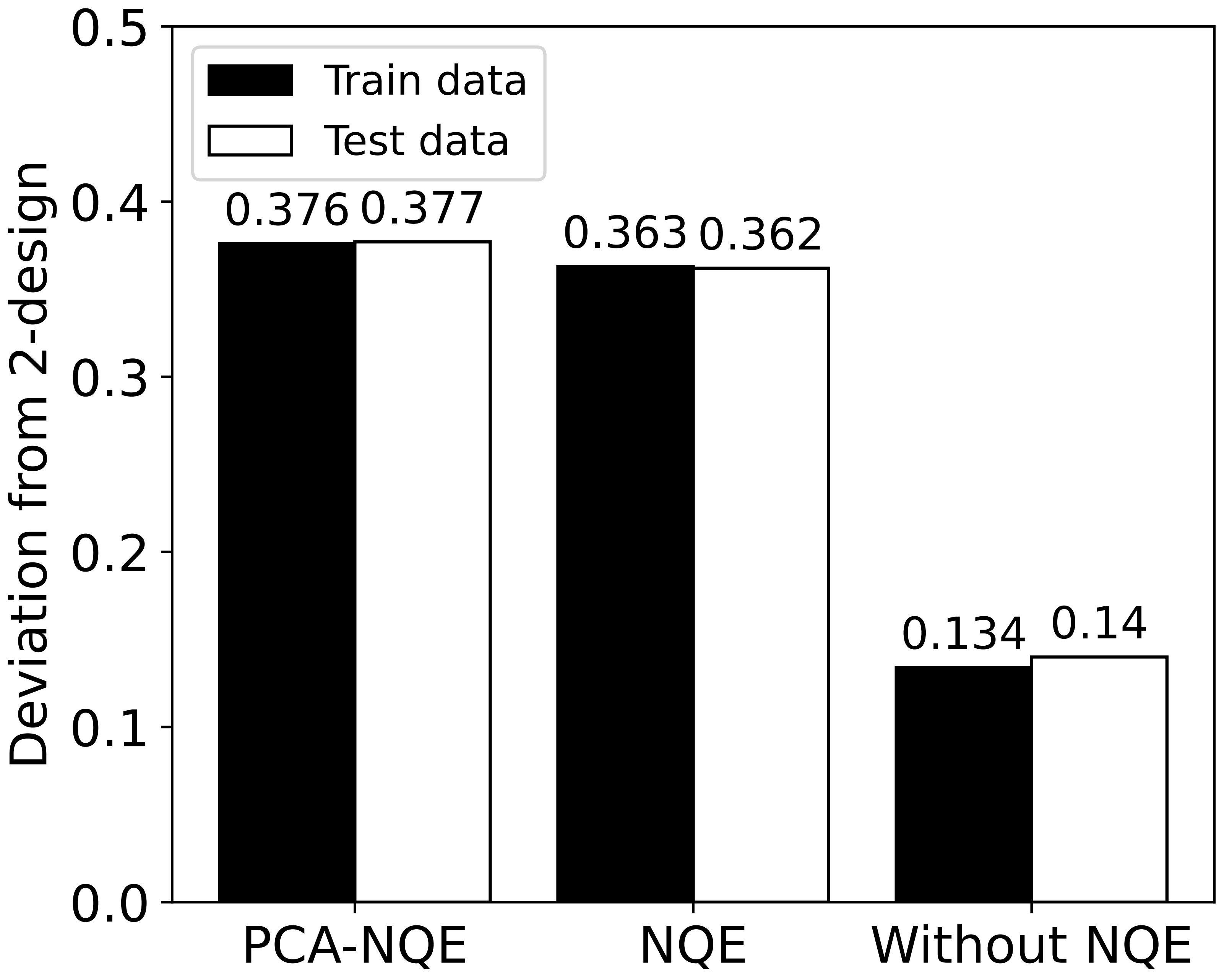


Expressibility & Trainability

1. Expressibility

Deviation from Unitary 2-Design,

$$A = \int_{\text{Haar}} (|\psi\rangle\langle\psi|)^{\otimes 2} d\psi - \int_{\mathscr{E}} (|\phi\rangle\langle\phi|)^{\otimes 2} d\phi$$

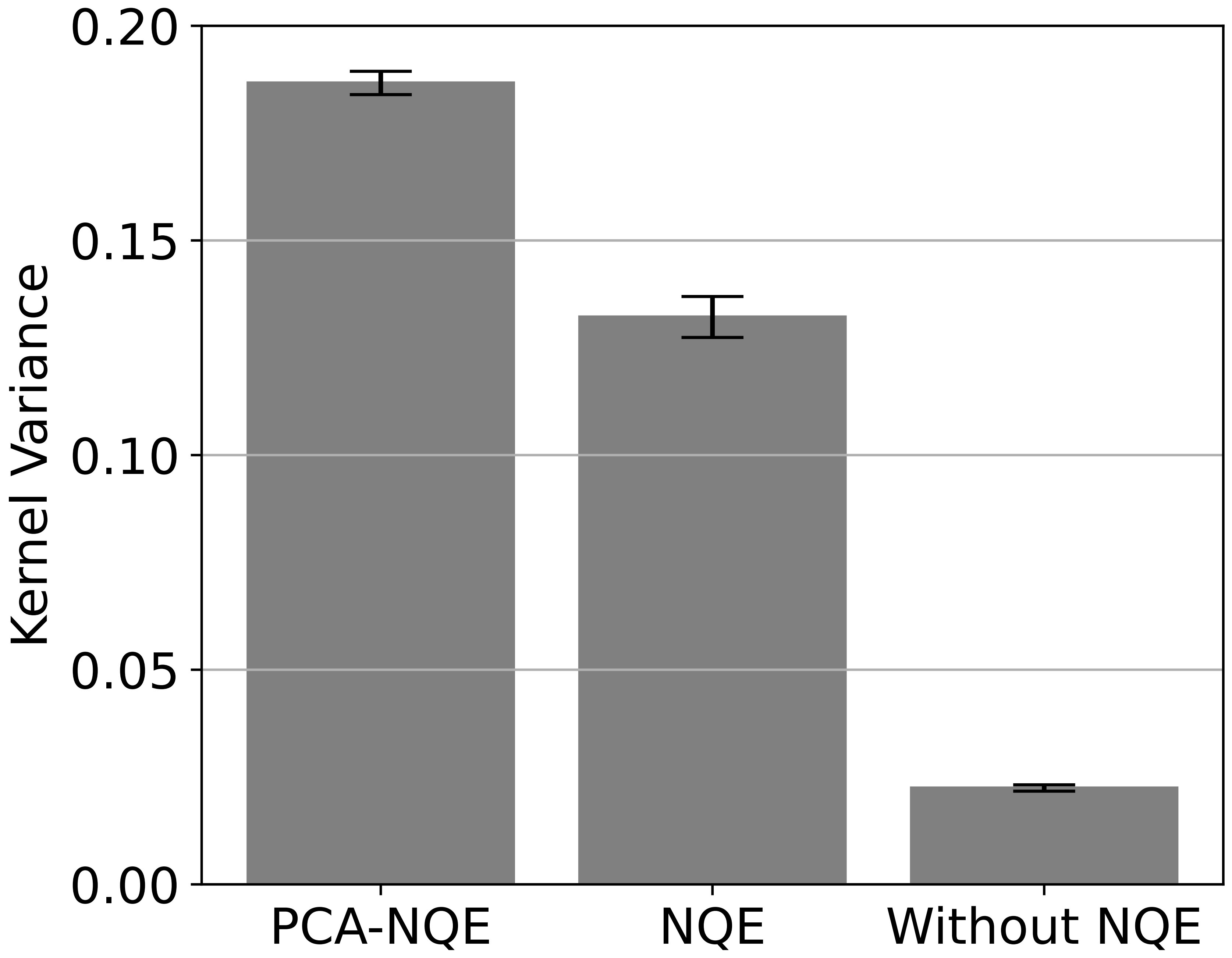


(a)

2. Variance of Kernel Elements

By Chebyshev's Inequality,

$$\Pr\left[\left|K^{Q}_{i,j} - \mathbb{E}\left[K^{Q}_{i,j}\right]\right| \geq \delta\right] \leq \frac{\mathrm{Var}\left[K^{Q}_{i,j}\right]}{\delta^2}$$



(b)

Summary

- Emphasize the importance of data separability of quantum embedding & introduce Neural Quantum Embedding

- Employing NQE improves many QML metrics including,
  - lower training error
  - higher classification accuracy
  - robustness against noise
  - improved generalization
  - improved trainability

Thank You!

arXiv:2311.11412