nature portfolio

Peer Review File

Machine learning on quantum experimental data toward solving quantum many-body problems

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

The primary contribution of this paper lies in the implementation of machine learning algorithms on quantum experimental data, specifically addressing diverse quantum many-body problems. The implementation is executed on IBM's superconducting quantum computing hardware. The obtained results validate the scalability and effectiveness of classical machine learning algorithms in processing quantum experimental data.

While the results are intriguing, a greater technical advance may be necessary for publication in Nature Communications. To enhance the manuscript, the authors could elaborate on the distinctions between their work and previous theoretical and experimental studies in applying classical shadow and machine learning algorithms to addressing many-body problems.

Given the paper's emphasis on applying classical shadows on noisy hardware, it may be pertinent to reference existing results on error mitigation tailored for classical shadows on noisy quantum devices, for example, [1-6].

Moreover, in recent years, various variants of classical shadows and also more general ensembles (apart from $Haar(2)^{\otimes n}$ that this paper considers) have been proposed [7-11]. The paper might benefit from incorporating more recent techniques from these studies. Since nonlinear properties are measured, the method here could be compared with hybrid approaches, like [12].

Details of the experiment are presented in Supplementary Note 5. I believe that there is enough detail provided for the work to be reproduced.

Overall, the paper is well-written and would be of interest to the quantum computing community. However, to be recommended for publication, more references may need to be added and the comments herein would need to be addressed.

Specific comments:

- Abstract: Adding the word "only" to the sentence "this approach has been realized for solving restricted problems because of the prevalence of noise in current quantum computers" might improve its clarify. It should be emphasized that this approach has been realized for solving *only* restricted problems and not more general problems.

- Supp Fig S2b: The y-axis of the figure could benefit from additional markings.

- In Eq. SE6, the NOT symbol "¬" should be explained.

- Supp Note 3: "If measurement outcomes give different particle number" should be replaced with "If measurement outcomes give different particle numbers".

- Supp Note 4 First Page Line 20: the sentence "... even O spans the entire system" is unclear.

[1] Chen et al, PRX Quantum 2, 030348 (2021). https://doi.org/10.1103/PRXQuantum.2.030348

[2] Koh et al, Quantum 6, 776 (2022). https://doi.org/10.22331/q-2022-08-16-776

[3] Jnane et al, https://arxiv.org/abs/2305.04956

[4] Brieger et al, https://arxiv.org/abs/2310.19947

[5] Zhao et al, https://arxiv.org/abs/2310.03071

[6] Wu et al, https://arxiv.org/abs/2310.12726

[7] Huang et al, Phys. Rev. Lett. 127, 030503 (2021), https://doi.org/10.1103/PhysRevLett.127.030503

[8] Hu et al, Phys. Rev. Research 4, 013054 (2022). https://doi.org/10.1103/PhysRevResearch.4.013054

[9] Bu et al, https://arxiv.org/abs/2202.03272

[10] Akhar et al, Quantum 7, 1026 (2023). https://doi.org/10.22331/q-2023-06-01-1026

[11] Zhou et al, Quantum 7, 1044 (2023). https://doi.org/10.22331/q-2023-06-29-1044

[12] Zhou et al, https://arxiv.org/abs/2208.08416

Reviewer #3 (Remarks to the Author):

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

Paper: Machine learning on quantum experimental data toward solving quantum many-body problems

Summary

This is an experimental paper implementing classical machine learning (ML) algorithms from Refs. [18, 19] for predicting ground state properties and classifying quantum phases. One key difference from previous work is that the training data for the ML algorithms is generated from quantum experiments run on IBM hardware whereas Refs. [18, 19] generated training data from classical simulation. In this way, the work assesses the performance of the algorithms trained on noisy data and their applicability in practice. For the task of predicting ground state properties, the authors implemented algorithms from Refs. [18, 19] and applied them to the 1D nearest-neighbor random hopping system and the Su-Schrieffer-Heeger Hamiltonian. In both cases, they consider a system size of 12 qubits and used error mitigation techniques to reduce the error in the training data. For classifying quantum phases, they implemented the algorithm from Ref. [18] and applied it to the cases of distinguishing a symmetry protected topological (SPT) phase from a trivial phase and a topologically ordered phase from a trivial phase. They consider system sizes of 44 and 25 qubits, respectively. For both tasks, the ML algorithm predictions agree reasonably with the exact values.

Strengths

One of the strengths of this paper is that it experimentally demonstrates a promising application for quantum computers before fault tolerance is fully achieved. In particular, it shows that these classical ML algorithms can still be used when data is gathered from quantum experiments, whereas previous numerical results for these ML algorithms only utilized data from classical simulation, which can quickly become infeasible at larger system sizes. I think it also interesting that the work incorporates error mitigation in the quantum experiments, emphasizing that measurement error mitigation can help reduce noise in training data to help the ML algorithms predict better. Another strength of the work is that for classifying quantum phases, previous experimental work only considers SPT phases and smaller system sizes. In contrast, by leveraging classical ML techniques, this work expands to classify topological phases and system sizes of up to 44 qubits.

Weaknesses

However, I think there are many weaknesses in the paper. Notably, the algorithms in this work have already been implemented in Refs. [18, 19], and this work only introduces minor modifications of them, which brings the novelty of the work to question. Moreover, the numerical experiments in Refs. [18, 19] made the training data (generated from classical simulation) noisy to assess the algorithms' performance in the presence of errors as well, so the motivation of the work to verify the robustness of ML algorithms is not strong to me. It appears

that the new methodology introduced is gathering the training data from quantum experiments and incorporating quantum error mitigation into this process, which I acknowledge above is an interesting experiment to carry out, but nevertheless does not seem to be the emphasis of the paper as written. There are some further comments on this in the suggestions section.

Another weakness of the work pertains to the issue of scalability. Scalability is emphasized as one of the main motivations of the work (see e.g., line 20, line 52), as errors in quantum computations may build up at larger system sizes, detrimentally impacting the effectiveness of classical ML algorithms trained on this erroneous data. I agree with this concern. However, all experiments are only conducted for a single system size, which is also quite small for the case of predicting ground states. It is not clear to me how one can discuss scalability without considering multiple system sizes and analyzing how the prediction error scales with system size. There are noisy simulations in the Supplementary Information that serve a similar purpose, illustrating how the prediction error scales with single-qubit, two-qubit, and measurement errors. However, these are not discussed in the main text and are not discussed in much depth even in the Supplementary Information.

Suggestions and Questions

First, related to the weakness section above, I suggest changing the framing of this work, especially in the abstract but also throughout the main text. To me, the main contribution of this work is to generate training data from quantum experiments and run previously known classical ML algorithms on this data. However, throughout the manuscript, it is sometimes unclear that such algorithms were already known, e.g., line 14 in the abstract. Moreover, it needs to be stated clearly what the novelty of this work is, particularly with respect to the numerical experiments in Refs. [18, 19]. These papers have already conducted numerical experiments, albeit with training data generated from classical simulation, demonstrating that the algorithms accurately solve the problems of interest, even in the presence of noise in the training data. The authors should address what we learn from their experiments as opposed to these numerical experiments. I also think that drawing more attention to the error mitigation techniques and noisy simulations would be useful.

Line 41-42: “but it requires the conversion of quantum data into classical data through the measurement process”

- I suggest rewriting the latter part of this sentence since, combined with the transition to the next paragraph, it makes it sound like methods such as the classical shadow formalism do not require a quantum to classical conversion. This is incorrect because a classical shadow is a classical representation of a quantum state obtained via measurements.

Reviewer #2 (Remarks to the Author):

Lines 49-50: “strictly more powerful than solving problems using classical computers only”

- I suggest writing this a bit more precisely. The references given prove that learning algorithms are more powerful with respect to computational complexity. In other words, a learning algorithm can (computationally) efficiently perform these tasks that are computationally hard for non-ML algorithms, under standard complexity theoretic assumptions.

Lines 75-77: “The central idea of classical shadow…”

- The idea to focus on predicting expectation values rather than a full description of a quantum state was first introduced by shadow tomography (Aaronson 2018). The classical shadow formalism takes inspiration from this and aims to construct an efficient representation of a quantum state that can be used to predict such expectation values.

Lines 84-85: “where each U_i^(t) is sampled from the Haar measure over the unitary group”

- It is also relevant to cite Paini and Kalev 2019 (https://arxiv.org/pdf/1910.10543.pdf) for this version using Haar-random single-qubit rotations. I also suggest mentioning in the text that you utilize a slightly different form of classical shadow from the perhaps more widely known version with single-qubit random Clifford rotations. In relation to this, was there a particular reason why the Haar-random single-qubit rotation version was used? It would be nice to have this explained briefly as well.

Line 90: Could you expand on what you mean by “direct measurement”?

In the “Case 1: Predicting the properties of the ground state” section, a reference to Ref. [19] should be added, along with a short statement saying that this is the algorithm you implement with some minor modifications. Similarly, in the “Case 2: Classifying quantum phases” section, a reference to Ref. [18] should be added. In this way, readers can easily find an extended discussion of these algorithms.

Lines 139-140: “properties of other systems of interest can be estimated without performing additional experiments”

This paragraph seems to reference transfer learning, in which the pre-trained model can be used to predict results for other, related tasks. It would be instructive for the reader if you

could add a brief explanation as to why you would expect this to work for the Hamiltonians you consider.

Lines 186-188: “Utilizing the same experimental data under Z_2 \otimes Z_2 symmetry, we trained a ML model and attempted to distinguish between the SPT and trivial phases among a total of 40 test data points from H_{CI}”

- I am confused by this sentence, stemming from the clause “utilizing the same experimental data.” Do you mean that you did not generate new training data for classifying phases for the H_{CI} Hamiltonian and used data for the H_{ZXZ} model? But the latter part of the sentence seems to imply that you trained a new model?

Line 260-261: “compressive quantum transformations… could lower the dimensionality of the quantum state”

- Do you have a reference for such transformations?

Typo in Figure 2b: “Rotaion for parity measurement” should be “Rotation for parity measurement”

We thank the referees for their comments and suggestions. Detailed point-by-point responses follow below. Changes to the manuscript are highlighted in yellow.

Reviewer #1 (Remarks to the Author):

"The primary contribution of this paper lies in the implementation of machine learning algorithms on quantum experimental data, specifically addressing diverse quantum many-body problems. The implementation is executed on IBM's superconducting quantum computing hardware. The obtained results validate the scalability and effectiveness of classical machine learning algorithms in processing quantum experimental data."

We reply: We appreciate the reviewer for the concise summary of our research.

1. "While the results are intriguing, a greater technical advance may be necessary for publication in Nature Communications. To enhance the manuscript, the authors could elaborate on the distinctions between their work and previous theoretical and experimental studies in applying classical shadow and machine learning algorithms to addressing many-body problems."

We reply: We appreciate the reviewer's suggestion and have added the summary of the following content to the main text.

Comparisons with theoretical papers - Existing theoretical papers [Huang et al. *Science* **377**, eabk3333 (2022), Lewis et al. http://arxiv.org/abs/2301.13169 (2023)] utilize data obtained through tensor network algorithms on classical computers to proceed with classical ML. As the system size increases, especially in dimensions higher than two, classical methods encounter limitations. In such cases, it becomes helpful to obtain training data through quantum computers. Our research demonstrates experimentally that not only in 1D but also in 2D cases, data from quantum computers can be used for solving problems of interest in many-body physics.

Comparisons with experimental papers – Past experimental papers [Zhang et al. *Phys. Rev. Lett.* **127**, 200501 (2021)] that use classical shadows have primarily focused on their original purpose of measuring physical observables of a given quantum state. However, after presenting an algorithm with mathematical rigor that classical shadows can be used not only for predicting physical quantities but also as data for classical ML [Huang et al. *Science* **377**, eabk3333 (2022)], to the best of our knowledge, we are the first to conduct an ML task using classical shadow obtained from an actual quantum computer.

Comparing our work with past studies that have undertaken similar tasks, one of our experiments focused on distinguishing between SPT and trivial phases. This task has served as a benchmark in various QML papers and was originally introduced in [Cong et al. *Nat. Phys.* **15**, 1273–1278 (2019)], with an experimental study on this subject also being published in *Nature Communications* [Herrmann et al. *Nat Commun* **13**, 4144 (2022)]. That experimental paper aimed to classify SPT phases from a trivial phase in a 7-qubit system, whereas our paper demonstrates the feasibility of this classification in a

44-qubit system. Furthermore, we experimentally showed that it is possible to classify quantum phases under more general conditions where translational symmetry is broken, conditions under which classification using QCNN becomes challenging [Liu et al. *Phys. Rev. Lett.* **130**, 220603 (2023)].

Page 3 line 5 : … Here, we experimentally verify the applicability of the hybrid approach to problems of interest in many-body physics. Previous research[18, 19] obtained training data for ML from tensor network algorithms on classical computers. These approaches face limitations as system sizes increase, particularly beyond two dimensions, highlighting the utility of quantum computers for generating training data. Our experiments demonstrate that classical ML on data from quantum computers is effective not only in one-dimensional but also in two-dimensional many-body physics problems. Experimentally, prior works[17] have used classical shadows primarily for estimating observables of a given quantum state. However, building on the rigorous algorithm with various error-reducing procedures applied to the training data, we have successfully used classical shadows obtained from quantum computers for ML tasks, extending their applications beyond predicting physical observables for a given state.…

2. "Given the paper's emphasis on applying classical shadows on noisy hardware, it may be pertinent to reference existing results on error mitigation tailored for classical shadows on noisy quantum devices, for example, [1-6]. Moreover, in recent years, various variants of classical shadows and also more general ensembles (apart from $Haar(2)^{\otimes n}$ that this paper considers) have been proposed [7-11]. The paper might benefit from incorporating more recent techniques from these studies.

We reply: We appreciate the reviewer's suggestion of extensive references to recent related research. We have added citations and explanations in the appropriate places in the main text and the Supplementary Information.

Page 12 line 16 : …Additionally, training the ML model with experimental data obtained by applying various error mitigation techniques or others methodology suitable for classical shadows represents a promising avenue for further research[52-63].…

Added references in the main text

52. Chen, S., Yu, W., Zeng, P. & Flammia, S. T. Robust Shadow Estimation. PRX Quantum 2, 030348 (2021).

53. Koh, D. E. & Grewal, S. Classical Shadows With Noise. Quantum 6, 776 (2022).

54. Jnane, H., Steinberg, J., Cai, Z., Nguyen, H. C. & Koczor, B. Quantum Error Mitigated Classical Shadows. Preprint at http://arxiv.org/abs/2305.04956 (2023).

55. Brieger, R., Heinrich, M., Roth, I. & Kliesch, M. Stability of classical shadows under gate-dependent noise. Preprint at http://arxiv.org/abs/2310.19947 (2023).

56. Zhao, A. & Miyake, A. Group-theoretic error mitigation enabled by classical shadows and symmetries. Preprint at http://arxiv.org/abs/2310.03071 (2023).

57. Wu, B. & Koh, D. E. Error-mitigated fermionic classical shadows on noisy quantum devices. Preprint at http://arxiv.org/abs/2310.12726 (2023).

58. Huang, H.-Y., Kueng, R. & Preskill, J. Efficient estimation of Pauli observables by

derandomization. Phys. Rev. Lett. 127, 030503 (2021).

59. Hu, H.-Y. & You, Y.-Z. Hamiltonian-driven shadow tomography of quantum states. Phys. Rev. Research 4, 013054 (2022).

60. Bu, K., Koh, D. E., Garcia, R. J. & Jaffe, A. Classical shadows with Pauli-invariant unitary ensembles. npj Quantum Inf 10, 6 (2024).

61. Akhtar, A. A., Hu, H.-Y. & You, Y.-Z. Scalable and Flexible Classical Shadow Tomography with Tensor Networks. Quantum 7, 1026 (2023).

62. Zhou, Y. & Liu, Q. Performance analysis of multi-shot shadow estimation. Quantum 7, 1044 (2023).

63. Zhou, Y. & Liu, Z. A hybrid framework for estimating nonlinear functions of quantum states. Preprint at http://arxiv.org/abs/2208.08416 (2023).

3.Since nonlinear properties are measured, the method here could be compared with hybrid approaches, like [12]."

We reply: In our paper, we measured the purity of quantum states which is a typical nonlinear property. The most well-known method for measuring purity is a swap test [Buhrman et al *Physical Review Letters*, 87(16) (2001)], but obtaining accurate measurement results of the purity by the swap test is challenging due to errors caused by the controlled swap gates. Recently, various methods utilizing randomized measurements [Elben et al. *Phys. Rev. Lett.* **125**, 200501 (2020)] have been introduced. The proposed paper [12] merges two approaches to leverage their respective advantages. We anticipated that with slight improvements in the performance of current noisy quantum computers, allowing for a reduction in errors from the swap test, [12] could be applied to calculate nonlinear physical quantities. We have added an explanation of the relevant parts with citations in the Supplement Information. In our paper, we used MLE-QST [Smolin et al. *Phys. Rev. Lett.* **108**, 070502 (2012)], which is suitable for obtaining accurate values for relatively small system sizes.

Supplementary page 34 line 10 : ... In the experiment described in the main text, due to the relatively small number of qubits, the Maximum-likelihood estimator quantum state tomography (MLE-QST)[31] was used to obtain each RDM in the feature vector $\varphi(\rho)$. Unfortunately, when calculating $\mathrm{Tr}(\rho_A^2)$ using this method, exponential resources in samples and computational time are necessary as the size of A increases. To circumvent this issue, the swap test[33] can be utilized. Recently, a method[32] combining randomized measurement[7] and the swap test has been introduced. Implementing this to calculate the purity in larger regions of the A leads to promising subsequent research.....

Added reference in the Supplementary Information

32. Zhou, Y. & Liu, Z. A hybrid framework for estimating nonlinear functions of quantum states. Preprint at http://arxiv.org/abs/2208.08416 (2023).

4. "Details of the experiment are presented in Supplementary Note 5. I believe that there is enough detail provided for the work to be reproduced."

We reply: We appreciate the evaluation that the current manuscript provides sufficient experimental details.

5. "Overall, the paper is well-written and would be of interest to the quantum computing community. However, to be recommended for publication, more references may need to be added and the comments herein would need to be addressed."

We reply: We are grateful for the reviewer's generally positive feedback. We have incorporated various suggestions into the main text and the Supplementary Information and added more references.

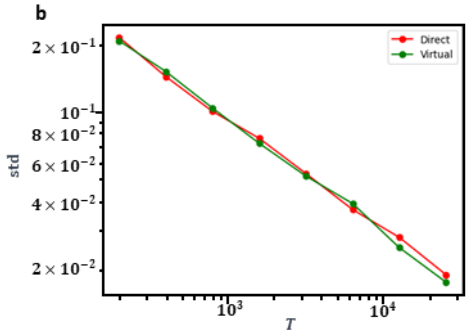
*6. Specific comments - Abstract: Adding the word "only" to the sentence "this approach has been realized for solving restricted problems because of the prevalence of noise in current quantum computers" might improve its clarify. It should be emphasized that this approach has been realized for solving *only* restricted problems and not more general problems.*

We reply: We agree with the reviewer's comments and have reflected this in the main text.

Page 1 line 12 : …this approach has been only realized for solving restricted problems…

7. Supp Fig S2b: The y-axis of the figure could benefit from additional markings.

We reply: We have added additional markings in Supp Fig. S2b to make it more noticeable.



8. In Eq. SE6, the NOT symbol "¬" should be explained.

We reply: We have added the definition of "¬" in (SE6).

Supplementary page 6 line 12 : …where $\neg A$ is a complement of A…

9. Supp Note 3: "If measurement outcomes give different particle number" should be replaced with "If measurement outcomes give different particle numbers".

We reply: Appreciating the detailed review of our paper, we have incorporated the suggested change.

Supplementary page 8 line 21 : … If measurement outcomes give different particle numbers …

10. Supp Note 4 First Page Line 20: the sentence "... even O spans the entire system" is unclear."

We reply: It means that O is a global operator that can't be decomposed into the sum of local observables. In many-body physics, when distinguishing quantum phases, various physical quantities exist. Some, like magnetization = $\sum_i Z_i$, consist of a sum of local observables, while others can be global, such as 1D string order parameters [Pollmann et al. *Phys. Rev. B* **86**, 125441 (2012)]. However, recent papers [Liu et al. *Phys. Rev. Lett.* **130**, 220603 (2023)] have reported that, in cases where the system's translational symmetry is broken, even global observables O can fail to distinguish SPT phases. We have added further explanation to that section.

Supplementary page 17 line 20 : ... even O acts non-trivially on the entire system which cannot be decomposed into the sum of local terms...

Reviewers #2 and #3 (Remarks to the Author):

Summary
This is an experimental paper implementing classical machine learning (ML) algorithms from Refs. [18, 19] for predicting ground state properties and classifying quantum phases. One key difference from previous work is that the training data for the ML algorithms is generated from quantum experiments run on IBM hardware whereas Refs. [18, 19] generated training data from classical simulation. In this way, the work assesses the performance of the algorithms trained on noisy data and their applicability in practice. For the task of predicting ground state properties, the authors implemented algorithms from Refs. [18, 19] and applied them to the 1D nearest-neighbor random hopping system and the Su-Schrieffer-Heeger Hamiltonian. In both cases, they consider a system size of 12 qubits and used error mitigation techniques to reduce the error in the training data. For classifying quantum phases, they implemented the algorithm from Ref. [18] and applied it to the cases of distinguishing a symmetry protected topological (SPT) phase from a trivial phase and a topologically ordered phase from a trivial phase. They consider system sizes of 44 and 25 qubits, respectively. For both tasks, the ML algorithm predictions agree reasonably with the exact values.

Strengths

One of the strengths of this paper is that it experimentally demonstrates a promising application for quantum computers before fault tolerance is fully achieved. In particular, it shows that these classical ML algorithms can still be used when data is gathered from quantum experiments, whereas previous numerical results for these ML algorithms only utilized data from classical simulation, which can quickly become infeasible at larger system sizes. I think it also interesting that the work incorporates error mitigation in the quantum experiments, emphasizing that measurement error mitigation can help reduce noise in training data to help the ML algorithms predict better. Another strength of the work is that for classifying quantum phases, previous experimental work only considers SPT phases and smaller system sizes. In contrast, by leveraging classical ML techniques, this work expands to classify topological phases and system sizes of up to 44 qubits.

We reply: We are grateful to the reviewers for providing a summary of our paper and a detailed explanation of the strengths of our paper.

Weaknesses

1. However, I think there are many weaknesses in the paper. Notably, the algorithms in this work have already been implemented in Refs. [18, 19], and this work only introduces minor modifications of them, which brings the novelty of the work to question.

We reply: Our research introduces novelties in the intersection of quantum computing and machine learning, leveraging the capabilities of noisy intermediate-scale quantum (NISQ) devices and error-reducing procedures to push the boundaries of what is

currently achievable with quantum technologies.

Machine learning through data from NISQ devices: Our work marks a significant difference from previous methods of obtaining data for ML, such as the density matrix renormalization group (DMRG) algorithm, which, although effective for one-dimensional systems, encounters substantial limitations in dimensions higher than two. [18, 19] relied on classical computers to implement DMRG for generating the data. In contrast, we have harnessed data from NISQ devices, showcasing that even with their inherent noise, they can provide valuable tools for ML. Our approach is particularly beneficial for studying systems beyond one dimension, where classical algorithms falter.

Various error reducing procedures for refined data generation: Running theoretically proposed algorithms on current noisy quantum computers without quantum error correction is not straightforward. For instance, a paper [Herrmann et al. *Nat Commun* **13**, 4144 (2022)] that experimentally implemented the theoretical paper on QCNN [Cong et al. *Nat. Phys.* **15**, 1273–1278 (2019)] circumvents the optimization process for gate parameters required in convolution and pooling layers, instead employing experimentally friendly classical post-processing. While this approach improves performance for specific problems, it might raise questions about its applicability to general problems. Therefore, we aimed to run scalable algorithms with provable guarantees, without significant modifications, and focused on addressing the experimental challenges that arise, rather than modifying the algorithm to be more experiment-friendly.

More specifically, in our paper, obtaining training data for 2D topologically ordered states was challenging due to hardware connectivity, making the conventional method of applying unitary gates [Satzinger et al. *Science* **374**, 1237–1241 (2021)] impractical due to the excessive overhead of swap gates. To overcome this, we utilized a recently introduced theoretical method [Lu et al. *PRX Quantum* **3**, 040337 (2022)] that uses measurement as a resource and leveraged the randomness of classical shadows to overcome technical challenges on adaptive gates after measurement, as explained in Supplementary Information Note 3. This demonstrated the feasibility of obtaining quantum data on quantum platforms where mid-circuit measurements are challenging.

To emphasize the novelty of our work, we have added the following aspect in the main text.

Page 3 line 5 : ... Here, we experimentally verify the applicability of the hybrid approach to problems of interest in many-body physics. Previous research[18, 19] obtained training data for ML from tensor network algorithms on classical computers. These approaches face limitations as system sizes increase, particularly beyond two dimensions, highlighting the utility of quantum computers for generating training data. Our experiments demonstrate that classical ML on data from quantum computers is effective not only in one-dimensional but also in two-dimensional many-body physics problems. Experimentally, prior works[17] have used classical shadows primarily for estimating observables of a given quantum state. However, building on the rigorous algorithm with various error-reducing procedures applied to the training data, we have successfully used classical shadows obtained from quantum computers for ML tasks, extending their applications beyond predicting physical observables for a given state.....

2. Moreover, the numerical experiments in Refs. [18, 19] made the training data (generated from classical simulation) noisy to assess the algorithms' performance in the presence of errors as well, so the motivation of the work to verify the robustness of ML algorithms is not strong to me.

We reply: We appreciate the constructive criticism of our paper. We refer to our reply to point #6 for a detailed response.

3. It appears that the new methodology introduced is gathering the training data from quantum experiments and incorporating quantum error mitigation into this process, which I acknowledge above is an interesting experiment to carry out, but nevertheless does not seem to be the emphasis of the paper as written. There are some further comments on this in the suggestions section.

We reply: We agree with the reviewer's criticism. As stated in point #1, we have emphasized that, unlike the previous research, we obtained data from an actual quantum computer and implemented various error mitigation techniques. Additionally, as noted in point #6, error mitigation not only reduces errors in data but also shows a beneficial trend in reducing the ML model's prediction error as the number of data increases. These details are included in the main text and Supplementary Information, as written in point #1 and point #6.

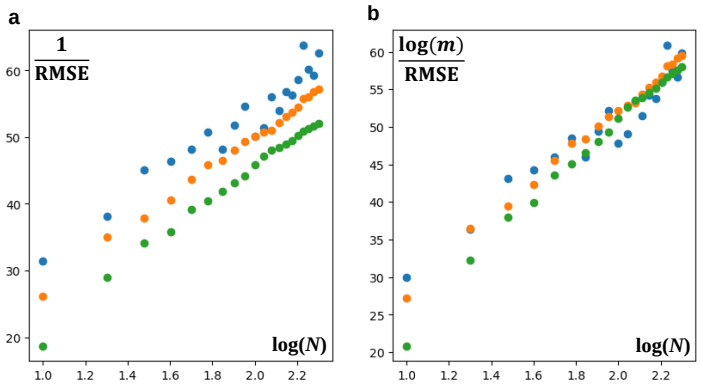
4. Another weakness of the work pertains to the issue of scalability. Scalability is emphasized as one of the main motivations of the work (see e.g., line 20, line 52), as errors in quantum computations may build up at larger system sizes, detrimentally impacting the effectiveness of classical ML algorithms trained on this erroneous data. I agree with this concern. However, all experiments are only conducted for a single system size, which is also quite small for the case of predicting ground states. It is not clear to me how one can discuss scalability without considering multiple system sizes and analyzing how the prediction error scales with system size. There are noisy simulations in the Supplementary Information that serve a similar purpose, illustrating how the prediction error scales with single-qubit, two-qubit, and measurement errors. However, these are not discussed in the main text and are not discussed in much depth even in the Supplementary Information.

We reply: First, we are grateful for the constructive criticism of the paper and have added the following discussions in proper places in the main text and Supplementary Information.

We separately discuss the scalability of each of the two tasks addressed in our paper. First, for the task of predicting properties of the ground state, we have shown in the main text (Fig. 2) the trend between the number of training data (N) and the model's prediction error. In theoretical papers, the scaling of N with the number of model parameters (m) for a fixed error, $N \sim m^{\mathcal{O}(1/\text{error})}$, has been obtained. However, in physical systems frequently

encountered, the number of parameters does not change as the system size increases but remains constant, for example, Fermi-Hubbard model has two parameters ($H = -t\Sigma_{i,\sigma=\uparrow,\downarrow}(c_{i,\sigma}^{\dagger} c_{i+1,\sigma} + \text{h.c.}) + U\Sigma_i n_{i\uparrow} n_{i\downarrow}$ where t is kinetic energy and U is onsite electron repulsion). In such cases, the primary interest is how increasing the number of data can lead to more accurate predictions. Therefore, in this regard, we have examined the scaling between N and prediction error in the main text while fixing the number of model parameters m.

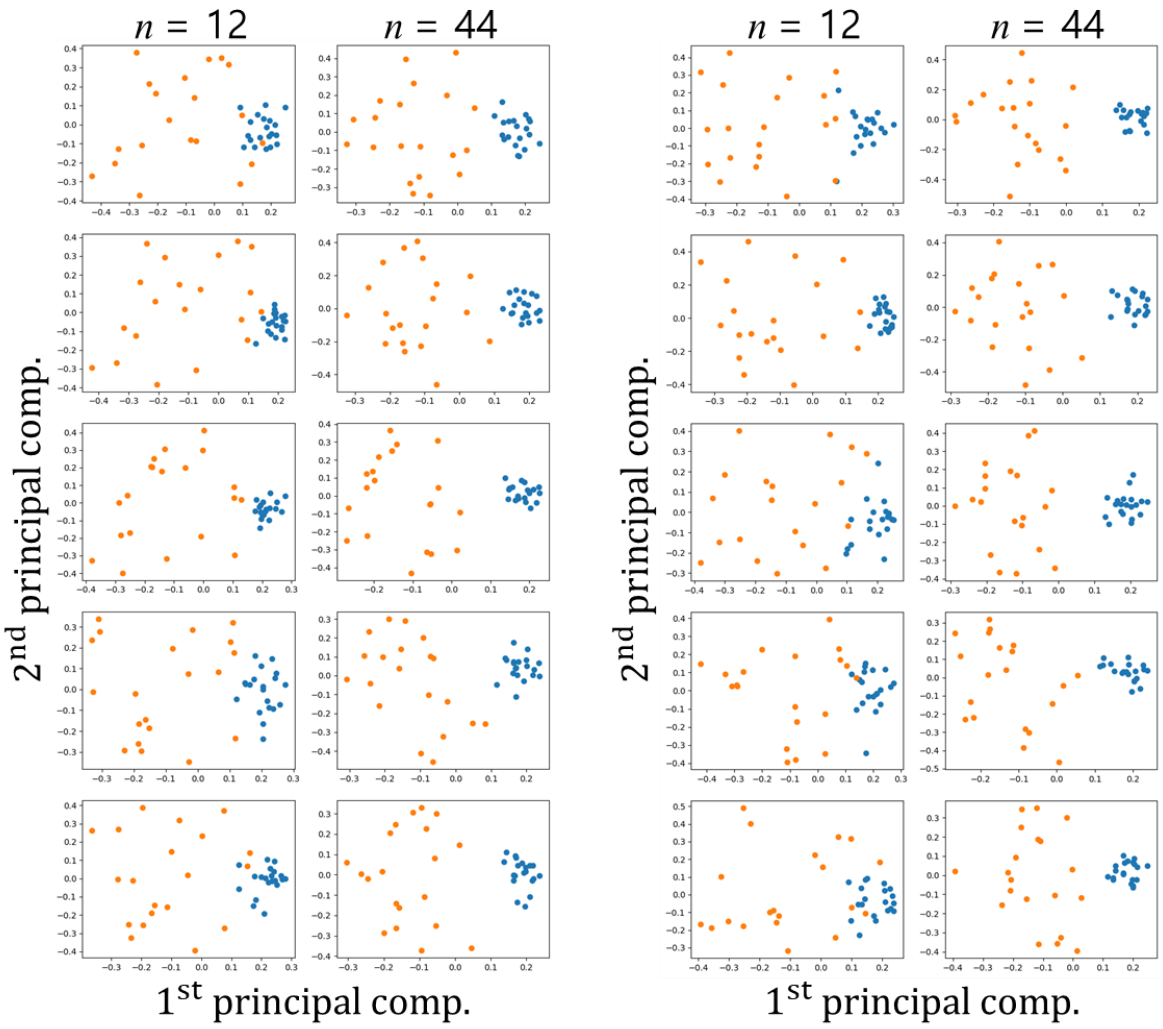
Although we showed the (N, error) scaling for the aforementioned reason, we agree with the reviewer's criticism and conducted additional experiments. By comparing cases with m = 9 (blue), 11 (orange), 13 (green), we obtained the following results. Note that the number of qubits (n) is equal to m + 1.



Revision Figure R1 | Comparison of the ML model's prediction error for different values of *m.* **a**, RMSE of the uncorrected ML model for different values of m. **b**, RMSE of the corrected ML model for different values of m.

As shown in the Fig. R1b, we observe a reasonable agreement in the scaling of $\log(m)/\text{RMSE} \sim \log(N)$ across different values of m, indicating that classical ML algorithms function effectively even as the system's number of parameters, m, increases in the presence of errors in the data (Ideally, the trends for different m would overlap, but there were slight deviations due to differences in errors of training data of each case of m).

Second, in the problem of classifying quantum phases, we conducted experiments with n = 44 qubits. The dependency of the algorithm's computational time on n appears when calculating the shadow kernel $k^{(\text{shadow})}(S_T(\rho_1), S_T(\rho_2))$ and is proportional to $\mathcal{O}(n)$. Therefore, even as the system size increases, the computational time only increases linearly. In addition, in our experiments, we found that when the number of layers of symmetric random unitary (d_{LU}) and T are the same, the larger the system size, the more clearly the quantum phase separates.



Revision Figure R2 | The data distribution for SPT with time reversal symmetry and trivial phase when T = 100, for n = 12 and n = 44. The x-axis corresponds to the first principal component, and the y-axis corresponds to the second principal component. Orange dots correspond to the trivial phase and blue dots to the SPT phase.

The phase boundary is relatively less clear for n = 12 compared to n = 44 (Fig. R2). In other words, as the system size increases, the task of separating quantum phases can be accomplished with a smaller T, and considering that the computational time of the algorithm is proportional to $\mathcal{O}(T^2)$, cases with a smaller n and larger T might actually take more computational time.

Page 6 line 24 : ... For a detailed comparison of the performance of the ML model across different values of m and between with and without error mitigation, refer to Supplementary Information Note 5.

Page 9 line 4 : ... The dependency of the algorithm's computational time on n appears when calculating the shadow kernel $k^{(\text{shadow})}(S_T(\rho_1), S_T(\rho_2))$ and is proportional to $\mathcal{O}(n)$. Therefore, even as the system size increases, the computational time only increases linearly. In addition, we found that when the number of layers of symmetric random unitary (d_{LU}) and T are the

same, the larger the system size, the more clearly the quantum phase separates. Further details on this are elaborated in Supplementary Information Note 5. …

Supplementary page 23 line 9, supplementary page 28 line 13: We have added the explanations and graphs described above in Supplementary Information Note 5 (5.1.7, 5.2.6).

Suggestions and Questions

5. First, related to the weakness section above, I suggest changing the framing of this work, especially in the abstract but also throughout the main text. To me, the main contribution of this work is to generate training data from quantum experiments and run previously known classical ML algorithms on this data. However, throughout the manuscript, it is sometimes unclear that such algorithms were already known, e.g., line 14 in the abstract.

We reply: We agree with the reviewer's comments and emphasize in our paper how our research differs from existing work by utilizing data obtained from quantum computers. We also mention that algorithms have already been developed in previous studies [Huang et al. *Science* **377**, eabk3333 (2022)].

Abstract: ... This enabled us to demonstrate the successful implementation of theoretically suggested classical ML algorithms ...

Page 5 line 16: … Among the many available algorithms, we used the previously studied kernel ridge regression (KRR)[18, 19] with minor modifications. ...

6. Moreover, it needs to be stated clearly what the novelty of this work is, particularly with respect to the numerical experiments in Refs. [18, 19]. These papers have already conducted numerical experiments, albeit with training data generated from classical simulation, demonstrating that the algorithms accurately solve the problems of interest, even in the presence of noise in the training data. The authors should address what we learn from their experiments as opposed to these numerical experiments. I also think that drawing more attention to the error mitigation techniques and noisy simulations would be useful.

We reply: In references [18, 19], a primary source of noise in training data could be identified as stemming from DMRG simulations. Typical errors in DMRG arise due to the truncation of bond dimensions in the matrix product state. However, these errors markedly differ from those encountered in quantum computers. For example, quantum computers may interact with the environment, leading pure states to evolve into mixed states—a phenomenon not occurred in DMRG simulations. Therefore, our research presents a notable distinction from [18, 19] by demonstrating that theoretically proposed ML algorithms can still be applied to data containing errors that might occur in real quantum computers.

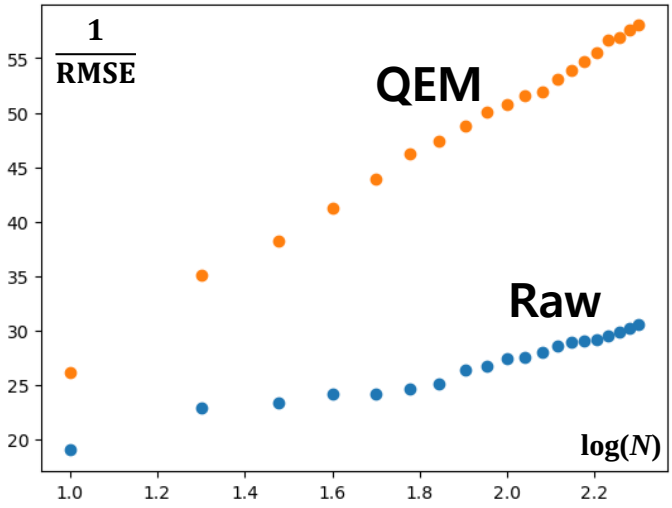
In the theoretical analysis of algorithm performance, where [19] considered a maximum error in the training data. In such cases, by utilizing more data in ML algorithms,

they show that it was possible to obtain an upper bound for the number of training data, even in cases where the data contained errors, regardless of their source. However, when the error in the training data and the model's target prediction error are similar, the upper bound becomes very loose, providing little insight into the actual model's prediction error. Denoting the target prediction error of the ML model as ε and the error in the training data as $\varepsilon_{\text{train}}$

$$\begin{aligned}\left(\mathrm{E}_x \| f_{\mathrm{ML}}(x) - \mathrm{Tr}(O\rho(x))\|^2\right)^{1/2} &\le \left(\mathrm{E}_{(x,y)} \| f_{\mathrm{ML}}(x) - y\|^2\right)^{1/2} + \left(\mathrm{E}_{(x,y)} \| \mathrm{Tr}(O\rho(x)) - y\|^2\right)^{1/2} \\ &= \left(\mathrm{E}_{(x,y)} \| f_{\mathrm{ML}}(x) - y\|^2\right)^{1/2} + \varepsilon_{\text{train}} \\ &\le \varepsilon \qquad \text{(RE1)}\end{aligned}$$

, then $\mathrm{E}_{(x,y)} \| f_{\mathrm{ML}}(x) - y\|^2 \le (\varepsilon - \varepsilon_{\text{train}})^2$ is required. To ensure this, $N = m^{O(1/(\varepsilon-\varepsilon_{\text{train}})^2)}$ needs. When substituting this with the experimental values $\varepsilon = 0.0172$ and $\varepsilon_{\text{train}} = 0.00959$, we get $1/(\varepsilon - \varepsilon_{\text{train}})^2 \simeq 17080$. In situations like our experiment, where the model's target prediction error is close to the error in the training data, the loose upper bound gives little insight about the actual prediction error. Thus, confirming that the ML algorithm remains effective even in the presence of errors encountered in real quantum computers is not assured by previous studies, marking the novelty of our research.

Despite the differences compared to [18, 19], we agree with the reviewer's suggestion that emphasizing error mitigation would be beneficial. Consequently, we have conducted an analysis comparing the outcomes of training with raw data versus error mitigated data and have included this discussion in the Supplementary Information.



Revision Figure R3 | Comparison of the ML model's prediction error with and without error mitigations.

We have observed that through error mitigation, not only can we obtain training data with lower errors, but as seen in the Fig. R3 above ($m = 11$), we can also achieve more favorable scaling of (N, error). Therefore, while the theoretical proof [18] aimed to demonstrate that the algorithm's sample and time complexity is polynomial on m, even if somewhat loose, our research study by investigating how various error mitigation techniques

practically benefit in case of quantum hardware error included.

We have added the explanation and graph mentioned above to Supplementary Information Note 5 (Section 5.1.7).

7. Line 41-42: "but it requires the conversion of quantum data into classical data through the measurement process" I suggest rewriting the latter part of this sentence since, combined with the transition to the next paragraph, it makes it sound like methods such as the classical shadow formalism do not require a quantum to classical conversion. This is incorrect because a classical shadow is a classical representation of a quantum state obtained via measurements.

We reply: We have deleted such part and rewritten to ensure that the meanings are more clearly conveyed.

Page 2 line 14 : ... Hybrid approaches in which classical computers are combined with quantum computers as to circumvent this issue have been introduced[14]. (deleted)

Research aimed at broadening the utility of the hybrid approach leads to quantum state learning methods such as classical shadows[15–17] which are succinct classical representation of the quantum state ...

8. Lines 49-50: "strictly more powerful than solving problems using classical computers only" I suggest writing this a bit more precisely. The references given prove that learning algorithms are more powerful with respect to computational complexity. In other words, a learning algorithm can (computationally) efficiently perform these tasks that are computationally hard for non-ML algorithms, under standard complexity theoretic assumptions.

We reply: In the main text, we have made changes that in terms of computational complexity, a classical computer with data from a quantum computer is more powerful than a classical computer alone.

Page 2 line 22 : ... Additionally, under widely believed complexity conjectures, the combined use of data from quantum computers and learning on classical computers can solve, in a computationally efficient manner, some problems that are challenging for non-ML algorithms relying solely on classical devices[18,22]...

9. Lines 75-77: "The central idea of classical shadow..."

The idea to focus on predicting expectation values rather than a full description of a quantum state was first introduced by shadow tomography (Aaronson 2018). The classical shadow formalism takes inspiration from this and aims to construct an efficient representation of a quantum state that can be used to predict such expectation values.

We reply: We are grateful for providing a comprehensive history of the relevant content. In the main text, we have added a reference to the foundational work by Aaronson (2018), while noting that Huang et al. (2020) have made it more practically applicable.

Page 4 line 7 : ... Despite many efforts to reduce the sample complexity for quantum state tomography (QST), exponential scaling of the sample complexity in terms of the system size (n) is unavoidable[30]. To circumvent the exponentially increasing sample complexity, research focusing on predicting physical quantities instead of the full description of a quantum state was introduced[50]. Subsequent studies presented the more experimentally friendly protocol of classical shadows[15]...

Added reference in the main text

50. Aaronson, S. Shadow Tomography of Quantum States. Preprint at http://arxiv.org/abs/1711.01053 (2018).

10. Lines 84-85: "where each U_i^(t) is sampled from the Haar measure over the unitary group"

It is also relevant to cite Paini and Kalev 2019 (https://arxiv.org/pdf/1910.10543.pdf) for this version using Haar-random single-qubit rotations. I also suggest mentioning in the text that you utilize a slightly different form of classical shadow from the perhaps more widely known version with single-qubit random Clifford rotations. In relation to this, was there a particular reason why the Haar-random single-qubit rotation version was used? It would be nice to have this explained briefly as well.

We reply: We have added the reference and a comment that the random unitary ensemble used in our experiments is different from widely used one. The reason we used Haar random single qubit gates (Haar(2)) rather than Clifford gates (Cl(2)) is because, in [Satzinger et al. *Science* **374**, 1237–1241 (2021)], it was reported that, statistically, Haar(2) performed better in certain cases compared to Cl(2).

Page 4 line 19 : … where each $U_i^{(t)}$ is sampled from the Haar measure over the unitary group $\mathbb{U}(2)$[51] instead of widely used Cl(2) (Clifford group on single qubit)…

Added reference in the main text

51. Paini, M. & Kalev, A. An approximate description of quantum states. Preprint at http://arxiv.org/abs/1910.10543 (2019).

11. Line 90: Could you expand on what you mean by "direct measurement"?

We reply: Direct measurement is a term contrasted with randomized measurement, referring to the case where the operator O to be measured is predetermined, and the measurement is performed in the basis that diagonalizes O. For example, if calculating the expectation value of observable $O = X_1Z_2X_3$ for a given quantum state ρ with 3 qubits, is applying a Hadamard gate (H) to the first and third qubits before measurement, and then performing a z-axis measurement on all qubits to estimate $\mathrm{Tr}(O\rho)$. To emphasize that

'direct measurement' is a term contrasted with 'randomized measurement,' we have added an explanation to that section.

Page 5 line 3 : …direct measurement was used for the regression, whereas classical shadow representation obtained by randomized measurement was employed for the classification…

12. In the "Case 1: Predicting the properties of the ground state" section, a reference to Ref. [19] should be added, along with a short statement saying that this is the algorithm you implement with some minor modifications. Similarly, in the "Case 2: Classifying quantum phases" section, a reference to Ref. [18] should be added. In this way, readers can easily find an extended discussion of these algorithms.

We reply: We have added citations to the relevant sections to facilitate easier reference for the readers.

Page 5 line 16 : …we used the previously studied kernel ridge regression (KRR)[18, 19] with minor modifications …

Page 7 line 14 : …principal component analysis (PCA) and a support vector machine (SVM) as classical ML algorithms1[18, 31] …

13. Lines 139-140: "properties of other systems of interest can be estimated without performing additional experiments"

This paragraph seems to reference transfer learning, in which the pre-trained model can be used to predict results for other, related tasks. It would be instructive for the reader if you could add a brief explanation as to why you would expect this to work for the Hamiltonians you consider.

We reply: In the text, we did not specifically have transfer learning in mind, but used the above statement to explain the advantage that, by utilizing a trained ML model, it is possible to make predictions about physically interesting systems later on without the need for additional physical experiments.

14. Lines 186-188: "Utilizing the same experimental data under Z_2 \otimes Z_2 symmetry, we trained a ML model and attempted to distinguish between the SPT and trivial phases among a total of 40 test data points from H_{CI}"

I am confused by this sentence, stemming from the clause "utilizing the same experimental data." Do you mean that you did not generate new training data for classifying phases for the H_{CI} Hamiltonian and used data for the H_{ZXZ} model? But the latter part of the sentence seems to imply that you trained a new model?

We reply: The ground state of H_{CI} includes a SPT phase with $\mathbb{Z}_2 \otimes \mathbb{Z}_2$ symmetry or a trivial phase that contains a simple product state, depending on the relative values of the

parameters in H_{CI}. In our experiments, by applying a random symmetric local unitary to the ground state of *H*zxz, we obtained data with $\mathbb{Z}_2 \otimes \mathbb{Z}_2$ symmetry, making it usable for predicting the quantum phase of H_{CI}'s ground state.

We computed the kernel matrix through the shadow kernel, then calculated two principal axes using PCA, and subsequently applied a classical ML algorithm. Therefore, even if the training data remains fixed, changing the test data alters the principal components of the kernel matrix, necessitating new training.

15. Line 260-261: "compressive quantum transformations… could lower the dimensionality of the quantum state"

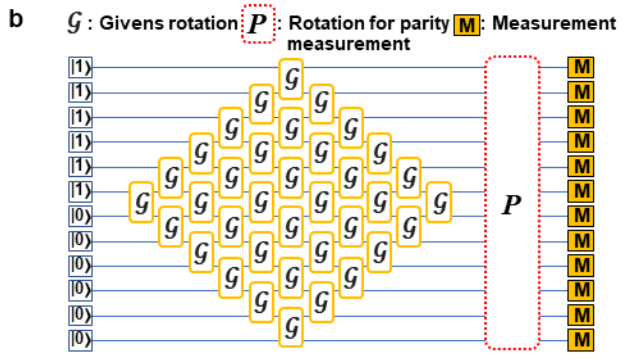
Do you have a reference for such transformations?

We reply: Any tensor renormalization group algorithms that can be implemented in a quantum circuit are possible, specifically, the method used in QCNN [Cong et al. *Nat. Phys.* **15**, 1273–1278 (2019)] involves repeating convolution and pooling layers to exponentially decrease the overall system size as the circuit depth increases. By reducing the system size and eliminating the local degrees of freedom of the quantum state, the training time is reduced proportionally to the decreased number of qubits. More importantly, it becomes possible to calculate physical quantities over a larger area with fewer measurements effectively, thus reducing the number of experiments (T) required to obtain the classical shadow of the quantum state ($S_T(\rho)$). Considering that the algorithm's computational time is $\mathcal{O}(nT^2)$, the training time of the ML model can be significantly reduced. We have added the reference to the corresponding line.

Page 12 line 12 : …compressive quantum transformations preserving an important feature of the state before measurements could lower the dimensionality of the quantum state[41]…

16. Typo in Figure 2b: "Rotaion for parity measurement" should be "Rotation for parity measurement"

We reply: We appreciate for the reviewer's careful proofreading. We have made the corrections and reflected them in the main text.



REVIEWERS' COMMENTS

Reviewer #1 (Remarks to the Author):

The authors have improved their manuscript by incorporating a comparison with previous works and adding more references to better contextualize their research. Specifically, they have compared their work with earlier theoretical studies that utilize data obtained through tensor network algorithms on classical computers, as well as experimental studies that use classical shadows to predict physical quantities associated with observables of a given quantum state.

Regarding its novelty, this work represents the first implementation of a machine learning task on a real quantum computer using classical shadows.

As previously mentioned, I believe that the paper is well written and that it would be of interest to the quantum computing community. I am satisfied with the changes made and would recommend it for acceptance if the minor points mentioned below are addressed.

There are a couple of minor grammatical errors and typographical errors that will need to be corrected. Please refer to the following.

On Line 43, there is a noun-verb agreement error in the sentence "... classical shadows which are succinct classical representation of quantum state".

On Line 285, in the sentence "Additionally, training the ML model with experimental data obtained by applying various error mitigation techniques or others methodology suitable for classical shadows represents a promising avenue for further research", the phrase "others methodology" is incorrect and should instead be replaced with "other methodology".

Reviewer #2 (Remarks to the Author):

I appreciate the authors' thorough response, especially the additional experiments run to address my comments. I have a few remaining comments.

The motivation and comparison to previous work are more clearly written in the updated version. Note that in Refs. [18, 19], while the numerical experiments were done using training data obtained from DMRG, the theoretical guarantees still hold for training data obtained from quantum computers (this is mentioned in both Refs. [18, 19]). This is an important distinction, as the current paper is not moving outside of the previously established theoretical framework in this regard. I recommend modifying the text to reflect this before acceptance.

I am still confused regarding the paragraph beginning at Line 149. From my understanding, the ML model is trained on data from the 1D NN random hopping system. At the same time, this paragraph describes that it is being used to predict properties of the SSH Hamiltonian without retraining the model on new data from the SSH model. The theoretical guarantees only apply when the model is trained and tested on data from the same model, with only the parameter values changing. I would appreciate more explanation in the text as to why this behavior is observed/why the authors would have expected it. Perhaps the explanation is similar to the response for phases (Point 14 in the previous review), which I would also appreciate being reflected in the text.

Reviewer #3 (Remarks to the Author):

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and

to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

We thank referees for their comments and suggestions. Detailed point-by-point responses follow below. Changes to the manuscript are highlighted in yellow.

Reviewer #1 (Remarks to the Author):

The authors have improved their manuscript by incorporating a comparison with previous works and adding more references to better contextualize their research. Specifically, they have compared their work with earlier theoretical studies that utilize data obtained through tensor network algorithms on classical computers, as well as experimental studies that use classical shadows to predict physical quantities associated with observables of a given quantum state.

Regarding its novelty, this work represents the first implementation of a machine learning task on a real quantum computer using classical shadows.

As previously mentioned, I believe that the paper is well written and that it would be of interest to the quantum computing community. I am satisfied with the changes made and would recommend it for acceptance if the minor points mentioned below are addressed.

There are a couple of minor grammatical errors and typographical errors that will need to be corrected. Please refer to the following.

We reply: We appreciate the reviewer's recommendation of our paper to Nature Communications.

On Line 43, there is a noun-verb agreement error in the sentence "... classical shadows which are succinct classical representation of quantum state".

We reply: We appreciate the reviewer's feedback on the grammar. We have updated the sentence.

Page 2 Line 18: ... classical shadow which is a succinct classical representation of a quantum state...

On Line 285, in the sentence "Additionally, training the ML model with experimental data obtained by applying various error mitigation techniques or others methodology suitable for classical shadows represents a promising avenue for further research", the phrase "others methodology" is incorrect and should instead be replaced with "other methodology".

We reply: We appreciate the careful reading. We have updated the sentence.

Page 12 Line 16: ... Additionally, training the ML model with experimental data obtained by applying various error mitigation techniques or other methodology suitable for ...

Reviewer #2 (Remarks to the Author):

I appreciate the authors′ thorough response, especially the additional experiments run to address my comments. I have a few remaining comments.

The motivation and comparison to previous work are more clearly written in the updated version. Note that in Refs. [18, 19], while the numerical experiments were done using training data obtained from DMRG, the theoretical guarantees still hold for training data obtained from quantum computers (this is mentioned in both Refs. [18, 19]). This is an important distinction, as the current paper is not moving outside of the previously established theoretical framework in this regard. I recommend modifying the text to reflect this before acceptance

We reply: We appreciate the reviewer's recognition of our previous response. To convey that our research confirms the applicability of theoretical results from [18, 19] to data obtained from actual quantum computers, rather than extending the theory itself, we included the following in the text.

Supplymentary Information page 19 Line 13: ... Our research demonstrated that the results predicted by theoretical papers are valid not only for numerically generated data but also for data from current noisy quantum computers with error mitigation.

I am still confused regarding the paragraph beginning at Line 149. From my understanding, the ML model is trained on data from the 1D NN random hopping system. At the same time, this paragraph describes that it is being used to predict properties of the SSH Hamiltonian without retraining the model on new data from the SSH model. The theoretical guarantees only apply when the model is trained and tested on data from the same model, with only the parameter values changing. I would appreciate more explanation in the text as to why this behavior is observed/why the authors would have expected it. Perhaps the explanation is similar to the response for phases (Point 14 in the previous review), which I would also appreciate being reflected in the text.

We reply: As in the proof from theoretical paper [18], the performance of the algorithm is valid only for data drawn from the same phase. However, in our actual experiments, we used a 1D random hopping system, and within the parameter range $x \in [0, 2]^m$, various quantum phases exist (including the two phases of the SSH model).

In our experimental results, the commonly used Gaussian kernel performed better than the Dirichlet kernel used in proof. A qualitative interpretation can be considered for this. The Gaussian kernel assigns larger weights only when the input parameters x and x' are close in L_2 distance to each other. On the other hand, as can be seen in many phase diagrams, except for the phase transition point, parameters x from different phases are far apart, causing the importance of the data points in different phases to decrease exponentially with the square of the distance due to the Gaussian kernel. Therefore, even

in training data that includes different phases, the prediction for a specific point x_{new} often relies on the training data in the vicinity of the parameter x_{new}, which approximately satisfied the assumption in [18,19]. We added the above content to the text as follows.

Supplymentary Information page 20 Line 14: ...In our experimental results, the commonly used Gaussian kernel outperformed the Dirichlet kernel, which was used in proof[10]. A qualitative interpretation may be considered. The Gaussian kernel assigns larger weights only when the input parameters x and x' are close in L_2 distance. Conversely, as observed in many phase diagrams, parameters x from different phases are usually far apart, except at the phase transition point. This spatial separation causes the importance of data points in different phases to decrease exponentially with the square of the distance, due to the Gaussian kernel. Therefore, even in training data that includes different phases, the prediction for a new point x_{new} often relies on the training data in the vicinity of x_{new}, which approximately satisfies the assumption in[10].

Reviewer #3 (Remarks to the Author):

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

We reply: We apreciate the reviewer's careful review.