Post-selection for QML and its Distribution across the Model

Gustav J L Jäger[1,2], Krzysztof Bieniasz[1], Hans-Martin Rieser[1]

1- Deutsches Zentrum für Luft- und Raumfahrt e.V. – Institut für KI-Sicherheit
Wilhelm-Runge-Straße 10, 89081 Ulm, Germany

2- Universität Ulm – Institut für Theoretische Physik and IQST,
Albert-Einstein-Allee 11, 89081 Ulm, Germany

May 4, 2026

1 Problem description and relevance

Next to efficiently optimizing a circuit covering an exponentially growing Hilbert space, quantum machine learning (QML) faces the difficult task of mapping classical data to the quantum computer such that unknown patterns can be identified with a unitary circuit [13]. Presently, this task is being addressed with research into encoding schemes [13] and methods based on data-reuploading [13, 4]. However, these methods are not applicable to quantum data, unless one combines multiple samples, which corresponds to data-reuploading. The core problem is that separation between two quantum states can not be increased using any quantum channel, which is known as the quantum data processing inequality [19]. Thus, any encoding that does not sufficiently separate two classes on the large Hilbert space has limited ability to classify them, especially when partially tracing to a smaller Hilbert space. The only way to increase the performance is to include non-linear pre- or post-processing steps, data-reuploading being the most prominent case. With post-selection we propose a more explainable approach, as compared to other non-linear operations, which directly identifies advantageous subspaces that offer better separation than the complete Hilbert space.

The use of post-selection in a machine learning context is motivated by the outperformance of classical tensor networks (CTNs) compared to quantum tensor networks on some problems [10]. Given some non-linearities in the encoding their optimal linear combination is usually not unitary. The core differences between CTNs and quantum circuits can be tied to two physical phenomena: First is entanglement, which relates to the bond dimension [3, 12]; Second is the amount of post-selection, see [11]. Commonly, CTNs reduce the Hilbert space without partial traces, quantum channels only employ partial traces [9, 18]. In previous work the authors proposed a hybrid approach between the two models [11], i.e. post-selection can be made trainable. By controlling a rotations on post-selected ancillas one can implement any diagonal matrix up to a global scaling factor on a quantum compute, a simple operation that is otherwise inaccessible on quantum computers, but easily implemented with a CTN. The core idea of implementing classical operations with post-selection is already known for other quantum algorithms, such as those solving partial differential equations [17, 15].

A big drawback for post-selection are the wasted shots on a quantum computer, which is why it is to be used sparingly. In this work, we aim to investigate how limited post-selection distributes itself inside a QML model. Understanding how this resource is distributed in models allows us to understand the structure of the selected subspace, which can be used as a stepping stone for the design of methods that treat the discarded subspaces. To this end, we use a hyperparameter from our previous work [11] that controls the amount of overall permitted post-selection. It controls the normalization of the output after post-selecting and corresponds to the number of shots kept after the post-selection. Thus, it allocates post-selection to the QML model during training, improving QML both for classical and quantum data.

2 Methodology

Let ρ_i be the density matrix of the encoded data i of the dataset. Following [11], we take the hyperparameter t that controls post-selection by inserting the normalization

$$\mathcal{N}_t[\rho_i] = \frac{\rho_i}{\max\left[\,\mathrm{tr}(\rho_i), t\right]} \tag{1}$$

into the loss. To ensure numerical stability of the logarithms in the cross entropy loss, during numerical experiments we add a depolarizing channel [19, 2]

$$\Delta_\lambda[\rho] = \lambda\rho + (1-\lambda)I. \tag{2}$$

However, $\lambda = 0$ simplifies theory development. Thus, the adapted crossed entropy loss L_t [8, 14] for a post-selected QML model reads

$$L_t = \sum_i \operatorname{tr}\left[\tau_i \log\left(\Delta_\lambda \circ \mathcal{N}_t \circ \Lambda[\rho_i]\right)\right], \tag{3}$$

with a completely positive (but not trace-preserving) map Λ. Partial post-selection which is not trace-preserving is implemented as follows: Arbitrary diagonal matrices $0 \le D \le I, D \ne 0$ are realized by controlling the rotation angle to ancillas that are fully post-selected [11]. Generally, the first diagonal element may be chosen 1. This procedure is equivalent to placing diagonal matrices on the partially post-selected location. For $t = 1$ no post-selection occurs, while for $t = 0$ no limits are placed on post-selection. While there are other choices for controlling the amount of normalization and thus post-selection, t is the most practical with its interpretation as a threshold, see Eq. (1).

Based on our simulation data we observed a trend, which we would like to investigate further. For the lack of more informal means, we define the hypothesis:

Hypothesis 1. *While partially post-selecting is permitted, qudits are either not post-selected at all or post-selected fully, except one state that is limited with the amount of permitted post-selection given by t.*

As a metric to investigate the hypothesis we identify all diagonal matrices D_j with their diagonal elements d_{jk} in the model that allow for partial post-selection and taking the number of $d_{jk} \in [0, 1]$ that are some distance ϵ from either 0 or 1. To this end, we take the counting function

$$h(d_{jk}) = \begin{cases} 1 & d_{jk} < \epsilon \\ 1 & 1 - d_{jk} < \epsilon \\ 1 & d_{jk} \text{ leaves trace invariant (up to } \varepsilon) \text{ for all data} \\ 0 & \text{else.} \end{cases} \tag{4}$$

Thus, we obtain the full score s to test Hyp.1 on a trained model

$$s = \frac{\sum_{jk} h(d_{jk})}{-1 + \sum_{jk} d_{jk}}. \tag{5}$$

Hyp. 1 is confirmed for scores $s \ge 1$.

When considering diagonal matrices acting on only a subset or individual qubits the diagonal matrices aggregate to an action on the entire Hilbert space via the tensor product. For instance, the partial post-selection of one qubit splits the entire Hilbert space in two subspaces that each get multiplied with respective factors. In this work we focus on the entries of the individual diagonal matrices and not their aggregated action because those may also be separated by layers of unitary operations or generally other quantum channels. Multi-qubit diagonal matrices allow for a higher precision, i.e. allow the selection of smaller subspaces.

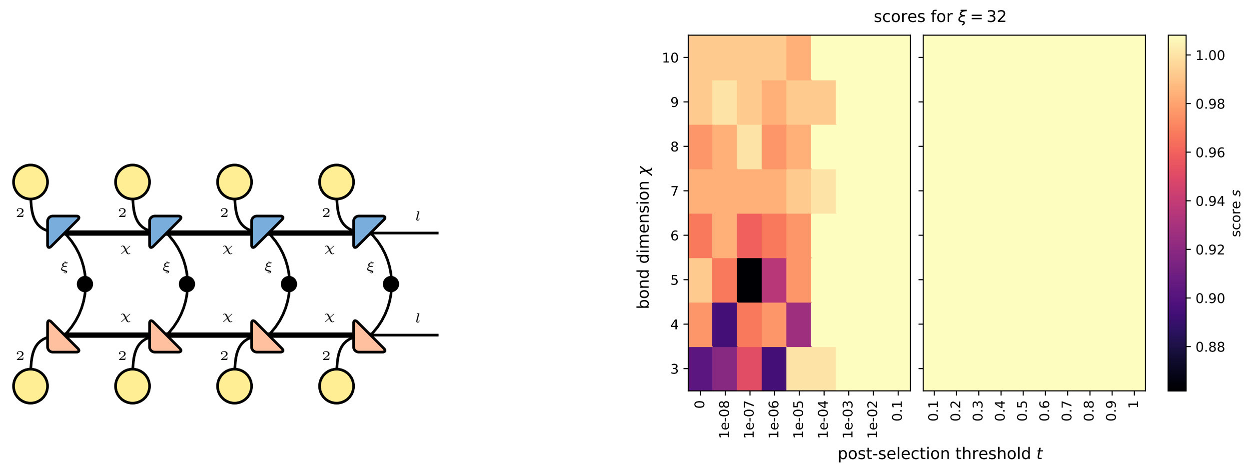
3 Practical demonstration

We investigate two models for different classification learning tasks. First is a model derived as a quantum classical hybrid via tensor networks [11]. It is a simple prototype that demonstrated the impact of post-selection on QML. However, in practical settings it is quite inconvenient as a quantum model because it encompasses the classical case as well. Thus, the second model will use more established methods with a simple brick wall ansatz with inserted post-selection layers.

3.1 Tensor Network Model

The tensor network based model building on [16], is motivated by the underlying theoretical arguments [11] leading up to Eq. (1). The tensor network is displayed in Fig. 1a. It uses rotational encoding of the iris dataset [5, 1] that is processed by an isometric matrix product state (MPS) with variable bond dimension χ during training and with trainable reduction operators of dimension ξ. The overall channel corresponds to a completely positive map and is trace preserving if all reduction operators are chosen to be the identities, i.e. partial traces. Choosing all reduction operators to be rank 1, i.e. doing post-selection, enables the channel to represent any unnormallized classical MPS.

The results are obtained by training on the full training data, and a stable initialization obtained from the data, s.t. the results for each hyperparameter pair χ and t are obtained from a single run. With this approach we observe overfitting for small $t < 0.1$, while testing scores and accurarcies significantly improve at $t \approx 0.5$. From the results in Fig. 1b we observe that the hypothesis holds for a sufficiently large threshold $t > 10^{-3}$.



(a) TN based QML model with post-selection

(b) scores for different χ and t

Figure 1: Results for the TN model from theoretic motivation

3.2 Brick Wall Circuit

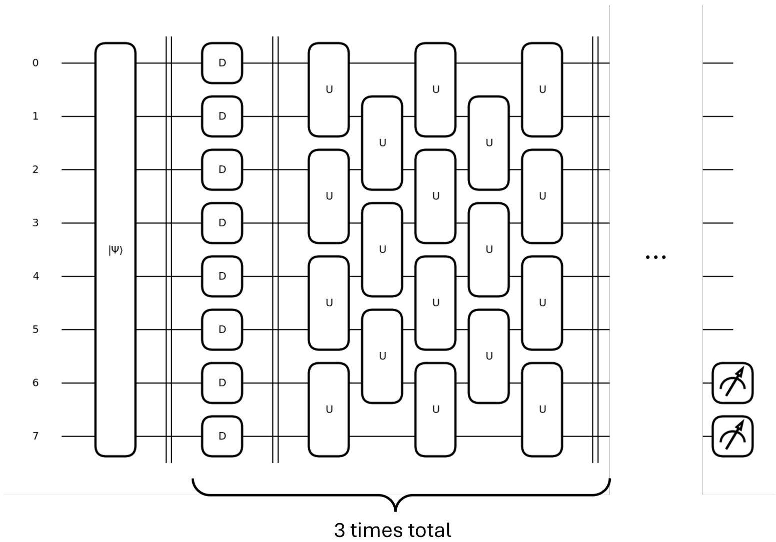
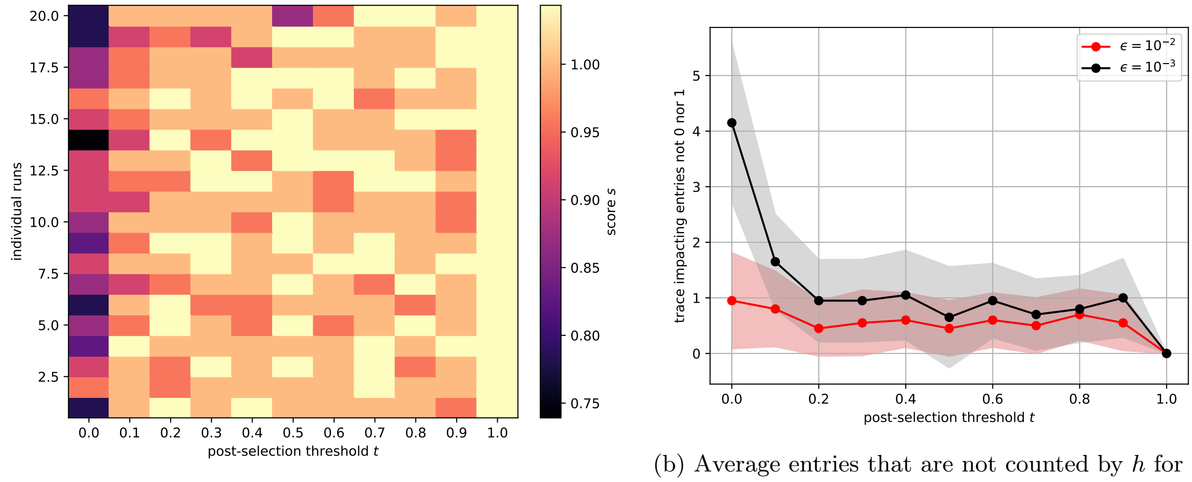


Figure 2: Quantum Circuit with amplitude encoding $|\Psi\rangle$, fully parameterized U, and $D_i = \begin{pmatrix} 1 & 0 \\ 0 & d_i \end{pmatrix}$ with $d_i \in [0, 1]$

To complement the theoretically motivated ansatz we probe a quantum circuit more native to QML [13] composed of amplitude encoding and brick wall layers of fully parameterized unitaries separated by inserted trainable partial post-selection operators. It is trained on the first 4 labels of the MNIST1D dataset [7]. The setup is batch trained with 10-fold cross validation with respective random intializations. Each run is repeated once with alternative choices of hyperparameters of the ADAM optimizer, resulting in 20 trained setups.

Small yet significant decrease on the training loss are observed with decreasing t, while changes in validations scores were not significant, suggesting inadequate placement of the partial post-selection operators. Confirming the threshold interpretation of t, the average traces across all the data and all 20 runs equal t. Thus, demonstrating that the trained circuits utilize the accessible post-selection. Similarly to Sec. 3.1, we observe that Hyp. 1 on average holds for larger values of $t > 0.1$. However, we observe that individual runs violate the Hypothesis. Like in Sec. 3.1 entries seem to stray from the distribution concentrated at 0 and 1. Thus, generally Hyp. 1 should be rejected and an amended version focusing on larger t should be considered. A possible explanation would be the limited post-selection forcing a more coarse grained use of partial post-selection which presents itself as the distribution in Hyp. 1. Instead of finer weighing of different subspaces, large sections of the Hilbert space are removed. This removed subspace rotates with a deacreasing threshold t, s.t. it has more overlap with the data resulting in a higher use of post-selection but no change in the entries of the operators imposing partial post-selection. Once most of the subspaces with negative classification performance are fully removed, further improvements can be made by careful weighing of the remaining Hilbert space, causing the observed change in entry distribution for $t < 0.1$ and in Sec. 3.1 for $t < 10^{-3}$.



(a) Individual scores for each run

(b) Average entries that are not counted by h for two choices of ϵ

Figure 3: Results for a standard brick wall ansatz with amplitude encoding.

4 Application Potential

From the data we observe that the original hypothesis does not generally hold. The observations for the larger thresholds coinciding with realistic choices do not extend to smaller thresholds, see Figs. 1b and 3b. Furthermore, individual runs in Fig. 3a seem to violate the hypothesis even for larger thresholds. There, instead of a single site we have a few sites that do not get counted by the counting function in Eq. (4). Thus, we reject Hyp. 1. However, we do observe a propensity of the diagonal entries to be mostly concentrated at the edge of the domain, i.e. either no post-selection or full post-selection is preferred on most sites.

We can identify application potential for post-selection in QML in the following points: Firstly, post-selection may be applied to improve a QML model or serve as a simple gauge on the further potential of non-linearities for the model. For quantum data it also offers a way to implement non-linearities without growing the Hilbert space to account for data-reuploading. Secondly, high potential lies in the explainability. The understanding that for sufficiently large t trainable partial post-selection leads to a pruning of the Hilbert space instead of a different weight distribution for the subspaces allows us to develop strategies that target and rectify the pruned subspace. A possible approach would be to replace post-selection with controlled operations that directly induce non-linearities, e.g. via data-reuploading, on the removed subspaces. Thus, post-selection might serve as a building block towards more advanced strategies. Ultimately, an encoding that is successively built up may be possible, which as a structured approach could address major issues in QML, such as barren plateaus, the choice of encoding, and the choice of circuit.

To this end, further understanding of the associated theory is needed. There also remain practical concerns how models with post-selection can be optimally trained, especially with regards to batching the data.

Acknowledgements

G.J received funding from the DLR (German Aerospace Center) through the Quantum Fellowship Program. Furthermore, the project was enabled by the DLR Quantencomputing-Initiative and the German federal ministry of Research, Technology and Space; `https://qci.dlr.de/projekte/qutenet`

References

[1] Edgar Anderson. “The Species Problem in Iris”. In: *Annals of the Missouri Botanical Garden* 23.3 (Sept. 1936), p. 457. ISSN: 0026-6493. DOI: 10.2307/2394164.

[2] Ingemar Bengtsson and Karol Życzkowski. *Geometry of Quantum States: An Introduction to Quantum Entanglement*. Cambridge University Press, Aug. 2017. ISBN: 9781107656147. DOI: 10.1017/9781139207010.

[3] Jacob C Bridgeman and Christopher T Chubb. "Hand-waving and interpretive dance: an introductory course on tensor networks". In: *Journal of Physics A: Mathematical and Theoretical* 50.22 (May 2017), p. 223001. ISSN: 1751-8121. DOI: 10.1088/1751-8121/aa6dc3.

[4] Yudong Cao, Gian Giacomo Guerreschi, and Alán Aspuru-Guzik. "Quantum Neuron: an elementary building block for machine learning on quantum computers". In: (Nov. 30, 2017). DOI: 10.48550/ARXIV.1711.11240. arXiv: 1711.11240 [quant-ph].

[5] R. A. Fisher. "The use of multiple measurements in taxonomic problems". In: *Annals of Eugenics* 7.2 (Sept. 1936), pp. 179–188. ISSN: 2050-1439. DOI: 10.1111/j.1469-1809.1936.tb02137.x.

[6] Johnnie Gray. "quimb: A python package for quantum information and many-body calculations". In: *Journal of Open Source Software* 3.29 (Sept. 2018), p. 819. ISSN: 2475-9066. DOI: 10.21105/joss.00819.

[7] Sam Greydanus and Dmitry Kobak. "Scaling Down Deep Learning with MNIST-1D". In: (2020). DOI: 10.48550/ARXIV.2011.14439.

[8] Philipp Grohs and Gitta Kutyniok. *Mathematical Aspects of Deep Learning*. Cambridge University Press, Nov. 2022. ISBN: 9781316516782. DOI: 10.1017/9781009025096.

[9] William Huggins et al. "Towards quantum machine learning with tensor networks". In: *Quantum Science and Technology* 4.2 (Jan. 2019), p. 024001. ISSN: 2058-9565. DOI: 10.1088/2058-9565/aaea94.

[10] Gustav Jäger, Martin B. Plenio, and Hans-Martin Rieser. "Quantum Tensor Network Learning with DMRG". In: *ESANN 2025 proceesdings*. ESANN 2025. Ciaco - i6doc.com, 2025, pp. 537–542. DOI: 10.14428/esann/2025.es2025-157.

[11] Gustav J L Jäger et al. "Entanglement is Half the Story: Post-Selection vs. Partial Traces". In: (Apr. 29, 2026). unpublished, submitted to Quantum Machine Intelligence. DOI: 10.21203/rs.3.rs-9561473/v1.

[12] Roman Orus. "A Practical Introduction to Tensor Networks: Matrix Product States and Projected Entangled Pair States". In: *Annals of Physics 349 (2014) 117-158* 349 (June 10, 2013), pp. 117–158. ISSN: 0003-4916. DOI: 10.1016/j.aop.2014.06.013. arXiv: 1306.2164 [cond-mat.str-el].

[13] Maria Schuld and Francesco Petruccione. *Machine Learning with Quantum Computers*. Springer International Publishing, 2021. ISBN: 9783030830984. DOI: 10.1007/978-3-030-83098-4.

[14] Zhou Shangnan and Yixu Wang. "Quantum Cross Entropy and Maximum Likelihood Principle". In: (Feb. 23, 2021). DOI: 10.48550/ARXIV.2102.11887. arXiv: 2102.11887 [quant-ph].

[15] Pia Siegl et al. *Tensor-Programmable Quantum Circuits for Solving Differential Equations*. 2025. DOI: 10.48550/ARXIV.2502.04425.

[16] E. Miles Stoudenmire and David J. Schwab. "Supervised Learning with Quantum-Inspired Tensor Networks". In: (2016). DOI: 10.48550/ARXIV.1605.05775.

[17] A. Termanova et al. *Tensor Quantum Programming*. 2024. DOI: 10.48550/ARXIV.2403.13486.

[18] Michael L. Wall et al. "Tensor-network discriminator architecture for classification of quantum data on quantum computers". In: *Phys. Rev. A* 105.6 (June 2022), p. 062439. ISSN: 2469-9934. DOI: 10.1103/physreva.105.062439.

[19] Mark M. Wilde. *Quantum Information Theory*. Cambridge University Press, Nov. 2016. ISBN: 9781316809976. DOI: 10.1017/9781316809976.