QuantumUQ: Uncertainty Quantification for Quantum Machine Learning with PennyLane and Qiskit

Ferhat Catak[1]

1 University of Stavanger, Norway

Summary

Quantum machine learning models can produce predictions that look confident even when they are unreliable. This can happen because quantum computers estimate outputs by sampling measurements ("shots"), because near-term devices are noisy, and because training quantum models can be unstable. **QuantumUQ** is an open-source Python library that helps researchers measure and compare predictive uncertainty in quantum machine learning workflows across two widely used ecosystems: **PennyLane** and **Qiskit**.

QuantumUQ provides tools to estimate shot-based variability, approximate model (epistemic) uncertainty using ensembles, profile prediction stability across different shot budgets, and evaluate reliability using calibration metrics such as reliability diagrams and expected calibration error (ECE). The library targets researchers who need reproducible, backend-agnostic uncertainty analysis for variational quantum classifiers and related QML models.

Statement of need

Uncertainty quantification (UQ) is a key element of trustworthy machine learning because it helps detect when a model's output should not be trusted. In quantum machine learning (QML), uncertainty has additional sources that are intrinsic to quantum computation: measurement outcomes are stochastic and expectation values are estimated with a finite number of shots, leading to sampling variance that depends directly on the measurement budget (Nielsen & Chuang, 2010). Furthermore, NISQ-era devices introduce hardware noise that perturbs measurement statistics (Preskill, 2018). Variational quantum circuits can also be difficult to train due to non-convex optimization and barren plateaus (McClean et al., 2018), and uncertainty-aware approaches are increasingly important for reliable evaluation (Kendall & Gal, 2017).

While QML frameworks provide powerful tools to construct and train circuits, they do not provide a unified, research-focused interface for uncertainty estimation, shot-budget profiling, and calibration analysis across backends. As a result, QML studies often report accuracy without systematically characterizing reliability. QuantumUQ fills this gap by providing a backend-agnostic UQ layer aimed at QML researchers who need consistent uncertainty metrics, reproducible experiments, and explicit resource–reliability trade-off analysis.

State of the field

PennyLane and Qiskit support circuit construction, simulators/hardware execution, and variational training. However, uncertainty-related evaluations are typically implemented ad hoc and are not standardized across frameworks. Classical uncertainty libraries and calibration toolkits are widely used in conventional deep learning, but they do not model shot-driven

39 uncertainty or measurement-budget trade-offs in QML. Calibration for probabilistic classifiers
40 is well studied in classical ML (Guo et al., 2017), yet QML papers rarely include systematic
41 calibration diagnostics.

42 QuantumUQ contributes a QML-specific layer that standardizes shot-based uncertainty
43 estimation, ensemble-based epistemic uncertainty (Lakshminarayanan et al., 2017), stability
44 profiling across shot budgets, and calibration metrics. It also provides adapters that make
45 these evaluations consistent across PennyLane and Qiskit. This design enables reproducible
46 research and fair comparisons across backends.

Software design

48 QuantumUQ is organized around (i) backend adapters, (ii) uncertainty methods, (iii) a
49 predictive distribution representation, and (iv) evaluation metrics.

50 Backend adapters normalize execution across PennyLane and Qiskit, enabling repeated circuit
51 evaluations at specified shot budgets. A `PredictiveDistribution` abstraction represents
52 repeated stochastic predictions and exposes common uncertainty summaries (mean, variance)
53 and reliability metrics (entropy and calibration).

54 QuantumUQ implements two core uncertainty estimators: shot-based estimators that quantify
55 measurement-induced variability, and ensemble-based estimators that approximate epistemic
56 uncertainty by training multiple independent models (Lakshminarayanan et al., 2017). The
57 library also includes stability profiling utilities that characterize how predictions change as shot
58 count increases, supporting principled selection of measurement budgets.

59 Design trade-offs are explicit: higher shot counts reduce sampling variance but increase runtime,
60 while ensemble methods increase cost linearly in the number of trained models. QuantumUQ
61 makes these trade-offs measurable and reproducible for QML workflows.

Example Outputs

63 QuantumUQ includes utilities for calibration analysis and measurement-budget profiling.
64 Figure 1 shows a reliability diagram generated using the library, illustrating mild overconfidence
65 in high-confidence regions. Figure 2 shows predictive entropy as a function of the number
66 of measurement shots for both PennyLane and Qiskit backends, demonstrating the expected
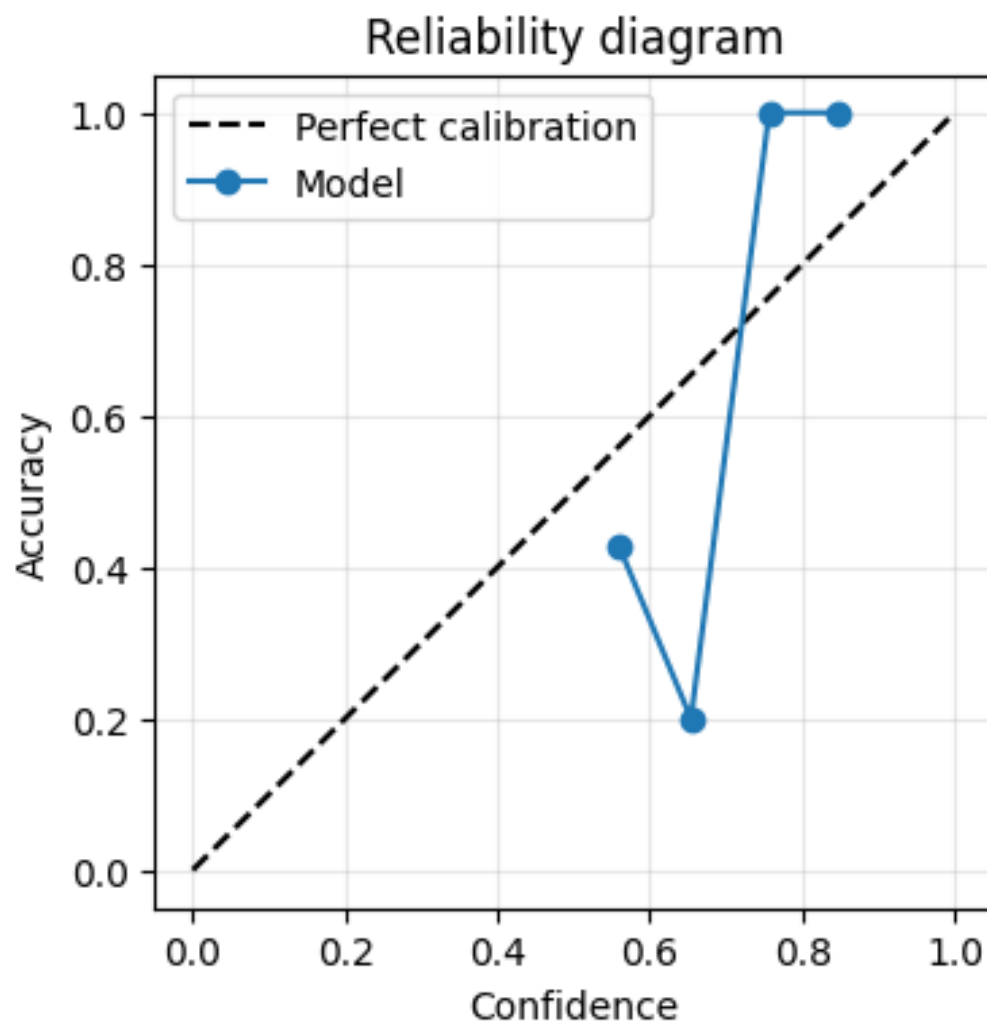67 decrease in uncertainty as the shot budget increases.

Figure 1: Reliability diagram illustrating calibration behavior of a variational quantum classifier.
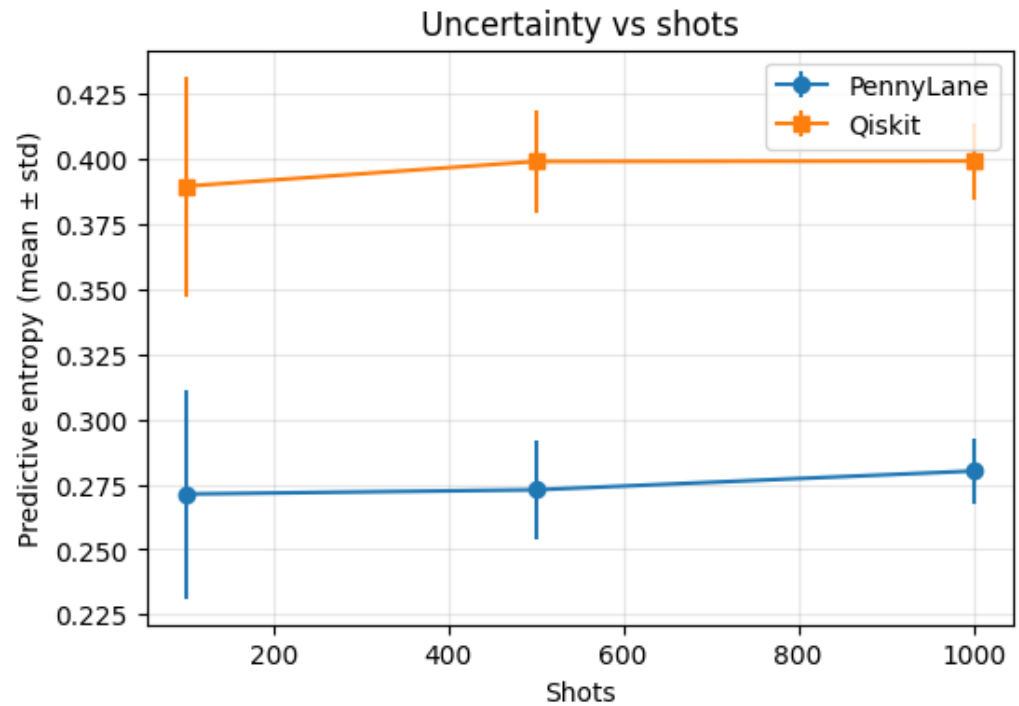
Figure 2: Predictive entropy as a function of measurement shots for PennyLane and Qiskit backends.

Research impact statement

QuantumUQ is distributed as an installable Python package and includes documentation and executable notebooks demonstrating usage with both PennyLane and Qiskit. The repository provides reproducible materials for generating calibration plots and uncertainty-versus-shots curves for variational quantum classifiers (Schuld et al., 2020). The package supports research questions central to reliability in QML: how many shots are required for stable predictions, where models are overconfident, and how uncertainty behavior differs across backends.

The software is designed to support near-term research impact by enabling consistent evaluation across QML frameworks and by providing reusable, reproducible building blocks for trustworthy quantum machine learning experiments.

AI usage disclosure

(Choose one.) - No generative AI tools were used in the software development, documentation, or paper writing. - Generative AI tools were used to assist with drafting and editing parts of the manuscript and/or documentation. All AI-assisted content was reviewed and edited by the author, and correctness was verified via unit tests and rerunning the provided notebooks end-to-end.

Acknowledgements

(Add funding, institutional support, compute resources, or project acknowledgements here.)

References

Guo, C., Pleiss, G., Sun, Y., & Weinberger, K. Q. (2017). On calibration of modern neural networks. *Proceedings of the 34th International Conference on Machine Learning - Volume 70*, 1321–1330.

Kendall, A., & Gal, Y. (2017). What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision? *arXiv e-Prints*, arXiv:1703.04977. https://doi.org/10.48550/arXiv.1703.04977

Lakshminarayanan, B., Pritzel, A., & Blundell, C. (2017). Simple and scalable predictive uncertainty estimation using deep ensembles. *Advances in Neural Information Processing Systems*, *30*.

McClean, J. R., Boixo, S., Smelyanskiy, V. N., Babbush, R., & Neven, H. (2018). Barren plateaus in quantum neural network training landscapes. *Nature Communications*, *9*(1), 4812.

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

Preskill, J. (2018). Quantum computing in the NISQ era and beyond. *Quantum*, *2*, 79.

Schuld, M., Bocharov, A., Svore, K. M., & Wiebe, N. (2020). Circuit-centric quantum classifiers. *Phys. Rev. A*, *101*, 032308. https://doi.org/10.1103/PhysRevA.101.032308