Fraunhofer-Institut für Kognitive Systeme IKS

Alona Sakhnenko, Julian Sikora, Jeanette Miriam Lorenz

Building Continuous Quantum-Classical Bayesian Neural Networks for a Classical Clinical Dataset

Introduction

- Some applications, e.g. medical AI, have strict reliability requirements where **trustworthiness of the confidence estimation** is crucial;

- **Bayesian learning** provides a well-established framework to train and analyse uncertainty-aware models;

- An approach should be capable to provide **(potential) advantage over classical benchmark**, **run on currently available devices** and be **easily scalable**.

In this work, we propose a **hybrid quantum-classical Bayesian Neural Network** (QCBNN) that is capable to uncertainty-aware classification of **ultrasound images**. It shows a **bigger gap in confidence of correctly and incorrectly identified samples** than its classical benchmark. Additionally, we perform a **systematic study of different quantum circuit architectures**.

Background

Bayesian learning

Bayesian Neural Networks (BNNs) are stochastic NN trained using **Bayesian inference**, which comprise of

1. a **stochastic model** of a chosen parametrization prior $p(\mathrm{w})$ and a prior of confidence $p(y|x, \mathrm{w})$, from which the posterior $p(\mathrm{w}|D)$ can be estimated,
2. a **functional model** $\Phi_{\mathrm{w}}(\mathrm{x})$, which in this case is a NN:

$$\mathrm{w} \sim p(\mathrm{w}|D)$$
$$\mathrm{y} = \Phi_{\mathrm{w}}(\mathrm{x}) + \epsilon$$

Probability of each class: $\hat{\mathrm{p}} = \frac{1}{N}\sum_{i=0}^{N} \Phi_{\mathrm{w}_i}(\mathrm{x})$

Our model

Quantum-classical BNN

=

Quantum circuit

+

Classical NN

Background

Variational Inference classically

Sampling directly from $p(\mathrm{w}|D)$ is **intractable**, therefore to train Bayesian models we need to approximate it.

For Variation Inference(VI) we select a family of tractable distributions $q_\theta(\mathrm{x})$ and tweak θ to minimize the **Kullback–Leibler (KL)** divergence between two distributions:

$$\mathbb{KL}[q_\theta(\mathrm{w}|D)||p(\mathrm{w}|D)] = \mathbb{E}_{\mathrm{w}\sim q_\theta(\mathrm{w}|D)}\left[\log \frac{q_\theta(\mathrm{w}|D)}{p(\mathrm{w}|D)}\right]$$

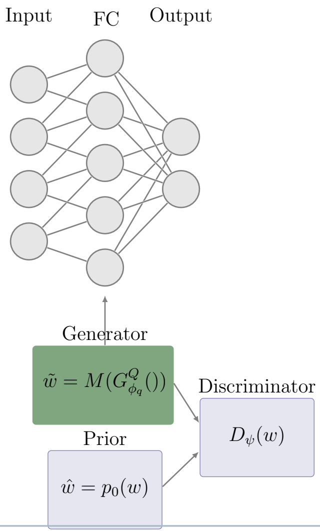
Variational Inference quantumly

Unfortunately, some parts of KL equation **cannot be measured** on a quantum computer.

[Benedetti et al., 2021] proposed an **adversarial learning loop** to go around this issue by introducing a discriminator model $d_\phi(\mathrm{w}, D)$:

$$\mathcal{L}_{KL}(\theta;\phi) = \mathbb{E}_{\mathrm{w}\sim q_\theta(\mathrm{w}|D)}[\mathrm{logit}(d_\phi(\mathrm{w}, D)) - \log p(D|\mathrm{w})]$$

Related work



Born Machines (BM) offer a way to model a distribution using a quantum state $|\psi(\theta, x)\rangle$. These models generate **bitstrings** with probabilities $q_\theta(w|x) = |\langle w|\psi(\theta, x)\rangle|^2$. The power of this method to model intractable distributions has been shown in [Coyle et al., 2020].

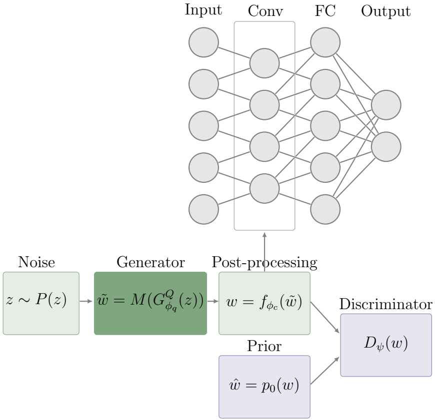
[Nikoloska and Simeone, 2022] proposed one of initial architectures idea for utilizing BMs in hybrid quantum-classical BNN setting.

- ✔ Showed utility on a toy example
- ✘ The weights of a NN are binary, which limits its applicability

Model

We assemble multiple architectural ideas that have proven to be viable in the literature to build a model with wider applicability range:

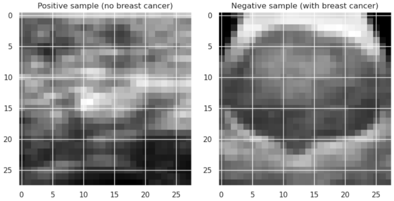
- **functional model** and its **application case**: (hybrid) Convolutional Neural Network on ultrasound images [Matic et al., 2022]
- **stochastic model**:
  - learning loop [Nikoloska and Simeone, 2022]
  - continues sampling technique [Romero and Aspuru-Guzik, 2019]
  - ensemble of 10 voters



Experimental setup

Dataset

We focus on BreastMNIST dataset, which consists of breast ultrasound 28×28 images that are classified into malignant and non-malignant categories. There are 780 samples.

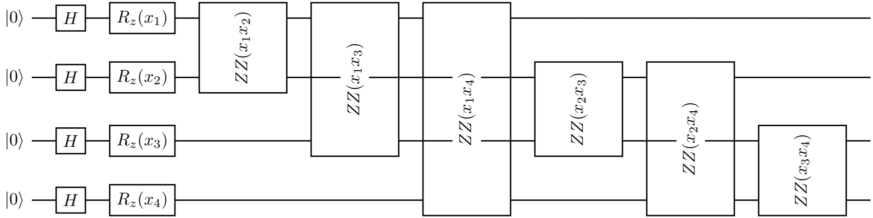


Environment

- **Classical part:** implemented with `PyTorch` and consists of a convolutional layer and a fully-connected layer.
- **Hybrid interface:** The weights of the convolutional layer are sampled from a distribution either produced by a classical NN (benchmark) or a PQC.
- **Quantum part:** PQC consist of a 4-qubits are implemented in `PennyLane`.

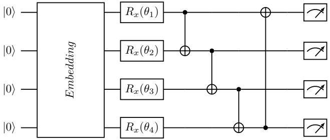
Quantum circuits: Embedding

We select an embedding strategy that showed the most promise in other works [Abbas et al., 2021, Matic et al., 2022] and our internal experiments, namely a **higher-order embedding** layer.

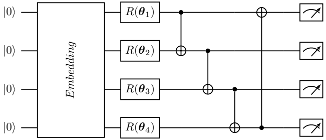


Quantum circuits: Calculation layers

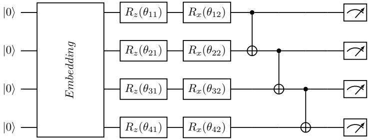
We start with architectural ideas exploited in the papers that we used to construct the model.



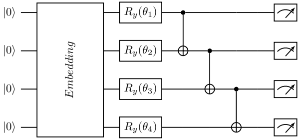
(a) *Matic I*



(b) *Matic II*

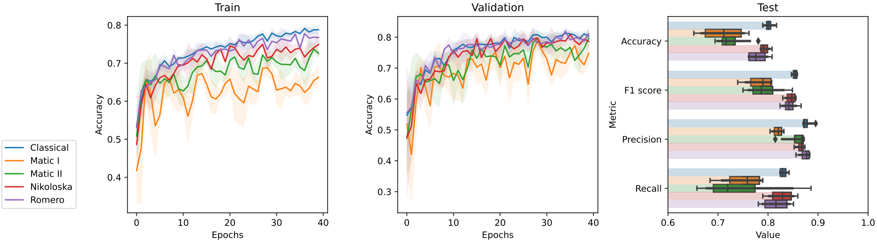


(c) *Nikoloska*

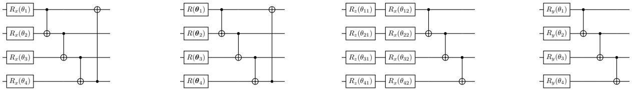


(d) *Romero*

Predictive performance

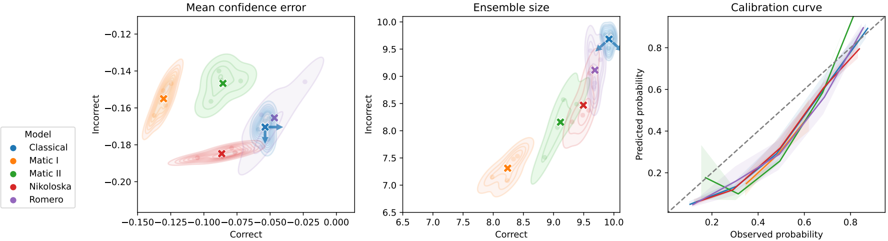


Reminder:

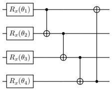


(a) *Matic I* (b) *Matic II* (c) *Nikoloska* (d) *Romero*

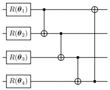
Uncertainty estimation



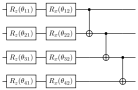
Reminder:



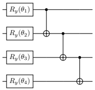
(a) *Matic I*



(b) *Matic II*



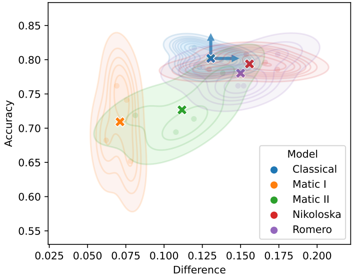
(c) *Nikoloska*



(d) *Romero*

Choosing architecture

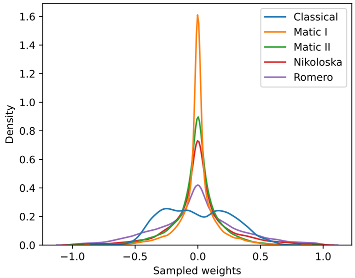
To find the best performing PQC we need to consider multiple behavioural parameters at once. To ease the analysis, we fuse the uncertainty metrics and compare it to the final accuracy.



$$\text{Difference} = \text{Confidence}_{\text{correct}} \times \text{Ensemble fraction}_{\text{correct}} - \text{Confidence}_{\text{incorrect}} \times \text{Ensemble fraction}_{\text{incorrect}}$$

From the above analysis we conclude that *Romero* architecture had one of the strongest performances.

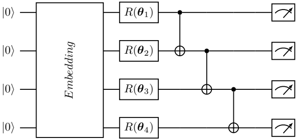
Weight distribution



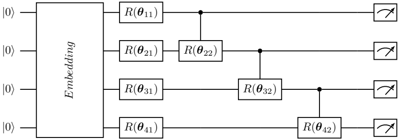
All of PQC's expectation values are **biased towards zero**. The architectures that had more **wide spread distributions** tend to have **stronger performances**.

Keeping this in mind, we proceed to tweaking the architectures in the following slides.

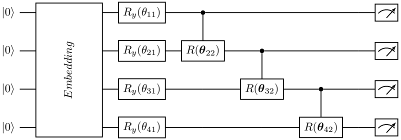
Quantum circuits



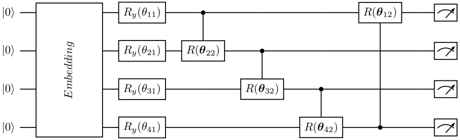
(a) *Circuit I*



(b) *Circuit II*

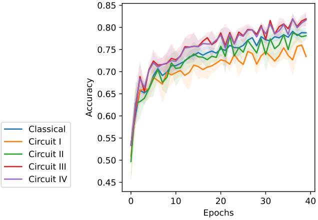


(c) *Circuit III*

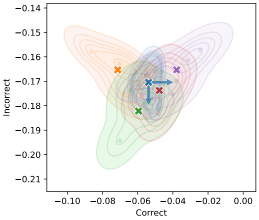


(d) *Circuit IV*

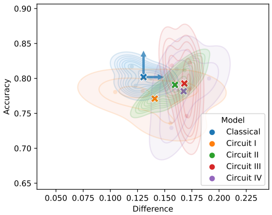
Predictive performance



(a) *Prediction*

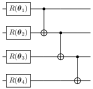


(b) *Uncertainty*

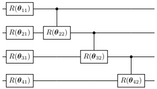


(c) *Combined*

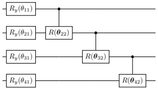
Reminder:



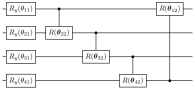
(a) *Circuit I*



(b) *Circuit II*

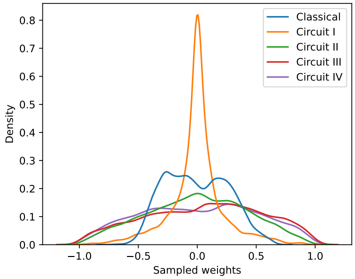


(c) *Circuit III*



(d) *Circuit IV*

Weight distribution



Circuit I's output is still biased towards $0.$, however, **changing the entanglement layer** allowed the distributions to **be wider** and **take more expressive forms**.

The models that performed at least the same as a classical benchmark also have at maximum the same density level, which aligns with a hypothesis that **BNN favour generators that provide diverse outputs**.

Discussion

- Results imply that by introducing **learnable parameters into the entangling layer** allows to achieve better performances for this task;
- A future research direction idea is to provide more **theoretical basis** that would allow to construct tailored PQCs in a less heuristic fashion, e.g. [Schuld et al., 2021];
- Here, we do not distinguish between **epistemic and aleatoric uncertainties**, e.g. [Nguyen and Chen, 2022] which is an interesting additional metric to consider next;
- Here, we modified a Born machine to allow for continuous weights, and hence **shifted the source of stochasticity from quantum to classical devices**. A further direction could be in developing something akin **Monte-Carlo Dropout** methods.

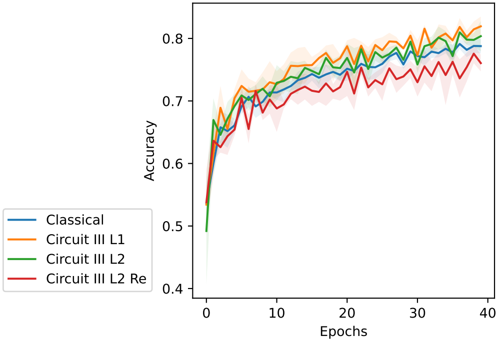
Conclusion

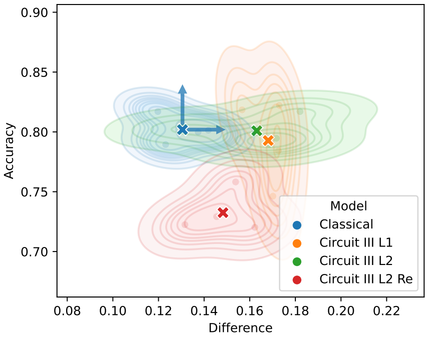
In this work, we

- introduce a Quantum-Classical Bayesian Neural Network (QCBNN) with continuous weights for uncertainty-aware classification of breast ultrasound scans (BreastMNIST);
- evaluate this model from the stand points of its predictive performance and uncertainty awareness;
- methodically test multiple PQC architectures in this scenario;
- determine that certain architectural features, such as trainable entanglement layers and rotation layers with less parameters, allow for better learning on this dataset;
- test these hypotheses by building custom PQCs for this task that boost the performance and even allow to slightly outperform the classical benchmark in terms of uncertainty awareness.

(Appendix) Saturating the performance

Can we improve the performance of *Circuit III* by increasing the number of layers? The layers can be added by simple repeating the computation layer, or adding both embedding and computation layers together to create a re-uploading model [Pérez-Salinas et al., 2020].





References I

- Abbas, A., Sutter, D., Zoufal, C., Lucchi, A., Figalli, A., and Woerner, S. (2021).
  The power of quantum neural networks.
  Nature Computational Science, 1(6):403–409.
- Benedetti, M., Coyle, B., Fiorentini, M., Lubasch, M., and Rosenkranz, M. (2021).
  Variational inference with a quantum computer.
  Physical Review Applied, 16(4).
- Coyle, B., Henderson, M., Le, J. C. J., Kumar, N., Paini, M., and Kashefi, E. (2020).
  Quantum versus classical generative modelling in finance.
- Matic, A., Monnet, M., Lorenz, J. M., Schachtner, B., and Messerer, T. (2022).
  Quantum-classical convolutional neural networks in radiological image classification.

References II

- Nguyen, N. and Chen, K.-C. (2022).
  Bayesian quantum neural networks.
  IEEE Access, 10:54110–54122.
- Nikoloska, I. and Simeone, O. (2022).
  Quantum-aided meta-learning for bayesian binary neural networks via born machines.
- Pérez-Salinas, A., Cervera-Lierta, A., Gil-Fuster, E., and Latorre, J. I. (2020).
  Data re-uploading for a universal quantum classifier.
  Quantum, 4:226.
- Romero, J. and Aspuru-Guzik, A. (2019).
  Variational quantum generators: Generative adversarial quantum machine learning for continuous distributions.

References III

Schuld, M., Sweke, R., and Meyer, J. J. (2021).
Effect of data encoding on the expressive power of variational quantum-machine-learning models.
Physical Review A, 103(3).