

Photonic quantum machine learning methods for Malware classification

Benjamin Stott¹, Grégoire Barrué², and Tony Quertier²

¹Quandela, France

²Orange Innovation, France

Introduction

Quantum Machine Learning (QML) is still a dynamic area of research, where lots of challenges are addressed, such as finding a concrete use-case exhibiting quantum advantage, using QML models to increase the speed of classical Machine Learning tasks, or even optimizing their consumption of energy. One of the main difficulties of QML methods is the phenomenon of Barren plateaus, where the gradients of the loss functions vanish exponentially fast [1]. Some works manage to identify patterns that can lead to Barren plateaus in variational quantum circuits [2], but very few models are proven to be Barren plateau safe. To circumvent this issue, one particular quantum algorithm seems to be a good candidate: the quantum convolutional neural networks [3, 4]. However, it has been shown that these models are quite always classically simulable [5]. One reason of this is also the fact that the datasets that we use or the use-cases where we want to apply these models are too simple. On the other hand, some people avoid the Barren plateau issue by using quantum kernel methods, which can describe quite every supervised learning task implying a quantum circuit [6]. For this kind of method, turning to Quantum Photonic circuits also allows to use shallow models and to control their expressivity with the input photon number [7].

In this paper, we want to apply all this knowledge about QML and photonic circuits to a concrete problem which is malware detection. This is an important cybersecurity topic, and we believe that QML can deal more easily with our datasets, which are much more complex than the usual datasets (MNIST, Iris, ...), than classical model which would need lots of samples and lots of features to train correctly.

1 Preprocessing and dataset

Malware detection is an important part of the cybersecurity area of research. We focus on Windows Portable Executable (PE) files, because that is the most represented category of files, and the most subject to attacks. We gathered them from datasets Bodmas [8] and PE Malware Machine Learning [9]. Our aim is to train an algorithm to detect if a certain file is malicious or not. In order to do so, we rely on a preprocessing called Ember [10] which allows to extract from the files more than 2,000 statistical and semantic statistical and semantic features. Once all these features are gathered, we performed feature selection in order to identify the ten most impactful features, and we computed a "malicious score" which gives a rate from 0 to 1 about how dangerous the file can be [11]. Thus, we finally deal with only 11 features per sample.

All of our quantum paradigms utilize identical preprocessing pipelines and evaluation protocols implemented through the MerLin framework, Qunadel's upcoming machine learning interface that abstracts quantum photonic circuit details into hyperparameters accessible to machine learning researchers. The 11-feature malware dataset (163k samples total) undergoes PCA reduction to 5-8 components suitable for quantum depending on the circuit size employed. Data partitioning uses a subset for train/test splits while the remaining dataset serves as an independent validation set, maintaining class distribution integrity. For image-based models, all images were resized to 32x32 before testing. Notably, a training size of 800 samples and a test size of 200 samples was used, taking the other 162,000 as a validation set. Tests with datasizes of 1,000 and 5,000 were also explored. A grid of models over training schemes, photon numbers, and circuit architectures are trained and compared in feature vector format and direct image-format for the QCNN model.

2 Algorithms and results

We evaluate three photonic quantum approaches on this benchmark. First, quantum kernel methods use photonic circuits as feature maps and train an SVM from the resulting kernel matrices. These experiments are particularly useful for studying how the representation changes with the number of modes and photons. Second, variational quantum neural networks (QNNs) train photonic circuits directly, either through measurement-to-label mappings or through a learned classical linear readout applied to the quantum out-

put distribution. Third, quantum convolutional neural networks (QCNNs) are applied to 32×32 image representations of the binary files, providing a complementary image-based quantum model.

As a classical reference, an RBF-kernel SVC trained on the same 1k-sample scale reaches approximately 94% validation accuracy. The best tabular photonic QNN reaches 92.18% accuracy, with macro F1 0.9216 and balanced accuracy 0.9219, approaching but not exceeding this classical kernel reference. On a 3-class malware-family task, the best QNN reaches 84.60% accuracy and macro F1 0.8393, showing that the approach is not restricted to binary detection. However, the QNN results are strongly dependent on hyperparameters such as feature count, number of modes, number of photons, measurement mapping, readout choice, and random seed; they do not yet show a clean monotonic scaling trend.

The photonic kernel experiments give a clearer scaling signal. Increasing the accessible Fock-space dimension through modes and photon number improves the best observed kernel configurations, which reach approximately 91% accuracy in the tested regimes, although the trend is not perfectly monotonic. The QCNN route is less competitive: on the 32×32 image representation with an 80/20 split, the best QCNN test accuracy reaches 72.15%, while fixed-train QCNN runs remain around 68% macro F1. Thus, QCNNs are included for completeness, but the tabular representation is currently more effective.

2.1 Discussion

The results demonstrate that the validation accuracy of both the kernels and the variational quantum neural network scale with the number of photons. The QCNN does not reach the same level of competitiveness. This suggests that the photon-scaling should be extended and that work should be done towards finding optimal configurations. Test accuracies reaching up to 93% and validation accuracies up to 92% were obtained for the various models tested. These results approach the accuracy of the fully classical model which obtained 94% on validation. Among the models that we tested, the kernel methods and QNN methods were most consistent for our cybersecurity use-case. Further work should be done on benchmarking the QNN on larger or unbalanced datasets and edge-cases to find a particular area where it overcomes the classical.

References

- [1] Michael Ragone, Bojko N. Bakalov, Frédéric Sauvage, Alexander F. Kemper, Carlos Ortiz Marrero, Martin Larocca, and M. Cerezo. A unified theory of barren plateaus for deep parametrized quantum circuits, 2023.
- [2] Martin Larocca, Supanut Thanasilp, Samson Wang, Kunal Sharma, Jacob Biamonte, Patrick J. Coles, Lukasz Cincio, Jarrod R. McClean, Zoë Holmes, and M. Cerezo. A review of barren plateaus in variational quantum computing, 2024.
- [3] Iris Cong, Soonwon Choi, and Mikhail D. Lukin. Quantum convolutional neural networks. *Nature Physics*, 15(12):1273–1278, aug 2019.
- [4] Arthur Pesah, M. Cerezo, Samson Wang, Tyler Volkoff, Andrew T. Sornborger, and Patrick J. Coles. Absence of barren plateaus in quantum convolutional neural networks. *Phys. Rev. X*, 11:041011, Oct 2021.
- [5] Pablo Bermejo, Paolo Braccia, Manuel S. Rudolph, Zoë Holmes, Lukasz Cincio, and M. Cerezo. Quantum convolutional neural networks are (effectively) classically simulable, 2024.
- [6] Maria Schuld. Supervised quantum machine learning models are kernel methods, 2021.
- [7] Beng Yee Gan, Daniel Leykam, and Dimitris G. Angelakis. Fock state-enhanced expressivity of quantum machine learning models. *EPJ Quantum Technology*, 9(1), June 2022.
- [8] Limin Yang, Arridhana Ciptadi, Ihar Laziuk, Ali Ahmadzadeh, and Gang Wang. BODMAS: An Open Dataset for Learning based Temporal Analysis of PE Malware. In *2021 IEEE Security and Privacy Workshops (SPW)*, pages 78–84, 2021.
- [9] Michael Lester. PE Malware Machine Learning Dataset.
- [10] Hyrum Anderson and Phil Roth. Ember: An open dataset for training static pe malware machine learning models. 04 2018.
- [11] Benjamin Marais. *Améliorations des outils de détection de malwares par analyse statique grâce à des mécanismes d’intelligence artificielle*. PhD thesis, Normandie, 2023.