A Data-Encoding Approach to Quantum Federated Learning: Experimenting with Cloud Challenges

Shiva Raj Pokhrel, Naman Yash, Jonathan Kua, Gang Li and Lei Pan

ABSTRACT

We examine the challenges of implementing Quantum Federated Learning (QFL) on cloud platforms, focusing on quantum complexities and platform constraints, and propose data-encoding-driven QFL, demonstrating its viability with genomic data sets.[1]

1 INTRODUCTION

Quantum computing's unparalleled computational prowess extends beyond classical capabilities, particularly in machine learning, revolutionizing sectors like healthcare, finance, and cybersecurity. Quantum Federated Learning (QFL) merges FL with quantum machine learning (QML) [1] on quantum networks, improving computational efficiency and model performance while preserving data privacy [2, 7] and security [4].

Our research centres on implementing a QFL algorithm on cloud-based quantum platforms, leveraging IBM's Qiskit library, and proposing a data-encoding-driven QFL approach. By comparing FL techniques, we aim to accelerate QFL adoption, enabling efficient quantum-enhanced training and learning models over the cloud while ensuring local privacy. Our proposed process transforms input data into quantum states, processes them using a parameterized quantum circuit, and iteratively optimizes the parameters for the desired outcomes.

Although QFL promises exponential speedups, quantum hardware development is still nascent, with major tech companies like IBM, Amazon, and Microsoft racing to provide cloud-based access to quantum processors. IBM leads with its 433-qubit quantum processor, while Amazon and Microsoft offer access to third-party processors like IONQ's Harmony and Rigetti's Ankaa-1. Despite challenges in software standardization and cloud-based access, IBM's Qiskit framework holds promise, albeit with challenges like limited developer resources. Despite obstacles, ongoing advances in quantum hardware and programming frameworks signal a bright future for QFL.

Qiskit's quantum simulators are indispensable tools in quantum computing, allowing developers to test algorithms on classical computers, and facilitating robust algorithm development and testing processes. These simulators, noted for their flexibility and accessibility, reduce the iterative refinement journey of algorithms, eliminating hardware access barriers while faithfully emulating actual quantum system behaviour. By accelerating rapid prototyping, they empower developers to iterate and refine algorithms efficiently, making embracing Qiskit's simulators a strategic imperative for advancing distributed quantum algorithm development.

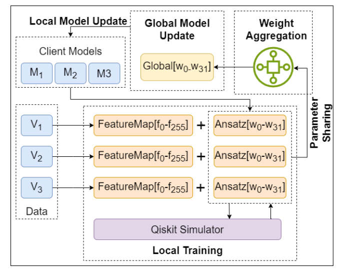

Figure 1: *Details of the proposed QFL framework.*

Advancements in Qiskit QML libraries are crucial for the progression of QFL [5]. Qiskit's existing library features implementations of key classifiers such as the Variational Quantum Classifier (VQC) and the Quantum Support Vector Regressor (QSVR), which serve as foundational elements for constructing QFL models. By providing ready-to-use modules, Qiskit streamlines the development cycle, enabling researchers and developers to focus on tailoring QFL models to specific application requirements.

2 PROPOSED QFL IMPLEMENTATION

In proposed QFL, the VQC extends to encode classical data into quantum states using the `RawFeatureVector()` feature map, with quantum circuits executed on various backends using the `Sampler()` component for flexibility in testing and running QML models. The ansatz, a parameterized quantum circuit, is iteratively optimized using the `COBYLA()` algorithm to improve classification accuracy by minimizing a loss function and optimizing QFL operationalization within the VQC architecture. Qiskit's quantum simulators and libraries enable us to develop and evaluate new QML and QFL algorithms easily, transitioning from traditional FL to decentralized model training while preserving data privacy. However, achieving QFL with quantum devices in each client poses significant challenges due to the absence of small-scale quantum devices, leading to client and global model training occurring primarily on the server in our QFL implementation utilizing IBM's quantum cloud platform and Qiskit's simulation capabilities, as depicted in Figure 1.

In our QFL, **data encoding** is crucial, seamlessly converting classical data into a quantum-compatible format, showcasing the potential of quantum facilities in handling complex genomic data [3]. Through innovative amplitude encoding, we streamline model training with our 200-feature dataset, utilizing only a fraction of the typically required qubits, especially advantageous in genomic data analysis.

Exploring *base, amplitude, and angle encoding* techniques, we find **amplitude encoding** as the most efficient, effectively representing vast amounts of information with minimal qubits. Our

[1]Authors are from IoT & SE Lab, School of IT, Deakin University, shiva.pokhrel@deakin.edu.au.

APNet 2024, August 03–04, 2024, Sydney, Australia

ACM ISBN 979-8-4007-1758-1/24/08
https://doi.org/10.1145/3663408.3665808

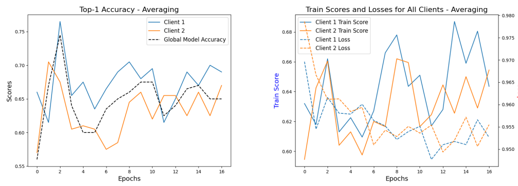

Figure 2: *With averaging, Left: Top-1 Accuracy over epochs for the global model and clients models; Right: Top-1 Accuracy and Loss over epochs for clients.*

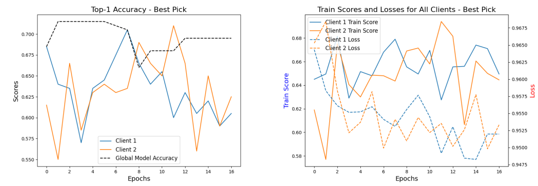

Figure 3: *With Best Pick, Left: Evolution of Top-1 Accuracy for the global model and individual clients. Right: training scores and losses for all clients.*

Algorithm ingeniously employs quantum circuits for data encoding and processing, offering a novel approach for genomic data analysis within a federated learning framework [6], leveraging the IBM Cloud platform for scalability and accessibility, paving the way for transformative advancements in genomic research with quantum-powered FL solutions.

In our setup, a server collaborates with a group of quantum computing clients to train a distributed Quantum Convolutional Neural Network model. Each client possesses its dataset of quantum states and labels, which forms the basis for individual model training. The objective is to optimize model performance, enabling accurate predictions based on each client's quantum state dataset, and advancing the intersection of quantum computing and FL towards transformative progress. Adhering to FL principles, we employ the Mean Squared Error (MSE) loss function locally via VQC. The proposed QFL process involves local training, parameter sharing, aggregation, global model update, and local model update stages, ensuring data privacy preservation while advancing collaborative quantum model training.

We redefine **global aggregation** by effectively managing client updates through three *weight aggregation schemes*. In Simple Averaging, we calculate the mean of parameters from individual client models to determine the new global model parameters. Weighted Averaging improves upon this by assigning weights based on the significance of each client's contribution, resulting in a weighted sum of individual client model updates. Finally, the Best Pick scheme adopts a selective approach, only incorporating updates from clients that meet a predefined performance criterion. This results in a weighted sum of selected client updates, prioritizing accuracy and reliability.

We employ **genomic datasets** [3] to train a decentralized QFL model, crucial to understanding genomic sequences, comprising labeled data points (X, Y) representing various genomic characteristics. We selected this dataset for its sufficient sample size for algorithm testing and the promising potential of genomics for future QFL applications. Demonstrated on the IBM Cloud platform, our algorithm utilizes multiple QML models, following the essential steps, including data preparation, model training, and aggregation of updates from multiple clients.

In our experiments, the QFL developed [5] yielded promising results, as shown in Figures 2 and 3, where the performance of the global model was primarily influenced by the highest-performing clients, with minimal or no contributions from the lowest performers, which requires further investigations and extension using techniques such as best arm identification [8].

3 CONCLUSION

We demonstrated QFL operationalization via Qiskit, showcasing how data encoding and weighted aggregation improved performance by capitalizing on high-performing clients while mitigating the influence of low-performing ones.

REFERENCES

[1] Jacob Biamonte et al. 2017. Quantum machine learning. *Nature* 549, 7671 (2017), 195–202.
[2] Mahdi Chehimi et al. 2024. Foundations of Quantum Federated Learning Over Classical and Quantum Networks. *IEEE Network* 38, 1 (2024), 124–130. https://doi.org/10.1109/MNET.2023.3327365
[3] Katarína Grešová et al. 2023. Genomic benchmarks: a collection of datasets for genomic sequence classification. *BMC Genomic Data* 24, 1 (2023), 25.
[4] Dev Gurung et al. 2023. Secure communication model for quantum federated learning: A proof of concept. *ICLR 2023* (2023).
[5] Shiva Raj Pokhrel et al. 2024. Quantum Federated Learning Experiments in the Cloud with Data Encoding. *arXiv preprint arXiv:2405.00909* (2024).
[6] Shiva Raj Pokhrel and Jinho Choi. 2020. Federated Learning With Blockchain for Autonomous Vehicles: Analysis and Design Challenges. *IEEE Transactions on Communications* 68, 8 (2020), 4734–4746. https://doi.org/10.1109/TCOMM.2020.2990686
[7] Shiva Raj Pokhrel and Mohammad Belayet Hossain. 2022. Data privacy of wireless charging vehicle to grid (V2G) networks with federated learning. *IEEE Transactions on Vehicular Technology* 71, 8 (2022), 9032–9037.
[8] Xuchuang Wang et al. 2023. Quantum Best Arm Identification. *SIGMETRICS Perform. Eval. Rev.* 51, 2 (oct 2023), 72–74. https://doi.org/10.1145/3626570.3626596