A Framework for Federated Quantum Machine Learning for Privacy-preserving Mobile Malware Detection

Fabio Martinelli[1], **Francesco Mercaldo**[2], **Antonella Santone**[2]

1. *Istituto di calcolo e reti ad alte prestazioni (ICAR), Consiglio Nazionale delle Ricerche (CNR), Rende (CS), Italy*
2. *University of Molise, Campobasso, Italy*

Abstract: We propose a framework for mobile malware detection considering quantum machine learning in combination with federated machine learning and considering explainability, by integrating the adoption of a way to visualise the model prediction.

Considering the inefficency of current antimalware in the detection of zero- day threats, reseachers are proposing the adoption of deep learning for malware detection, with particular regard to the mobile ecosistem. Nonetheless, there are several issues limiting the adoption of these methods in the real-world domain, the first one is the high presence of false positive, the second one is related to privacy and the third one is the lacsk of explainability. For these reasons, in this paper we propose a framework for mobile malware detection aimed to overcome these limits, in particular we propose the adoption of quantum machine learning [1], that demonstrated its supremacy with respect to classic deep learning mod- els, as for instance convolutional neural networks, in combination with federated machine learning [2], an emerging paradigm aimed to enable a set of clients to build a common model without share the data (only the updated weights) and considering explainability, by integrating the adoption of a way to visualise the model prediction.

Figure 1 shows the workflow of the proposal.

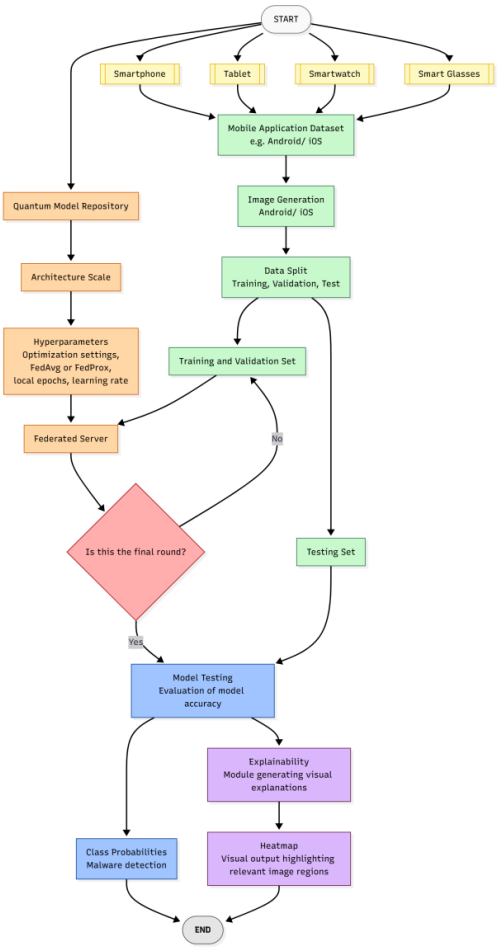

Fig. 1 The workflow of the proposal.

The workflow starts with decentralized client devices, such as smartphones and tablets, that collect mobile applications (e.g., Android/iOS). These data are transformed into images samples, which are then processed by a Quantum Model.

The dataset is divided into training, validation, and testing sets. During training, a federated server coordinates model updates from all clients. Each client performs local training on its own data using defined hyperparameters. The locally trained weights are periodically sent to the server, which aggregates them to update the global model until convergence is achieved.

Once training is complete, the global model is tested to assess its accuracy and ability to distinguish between benign and malicious applications. An explain- ability module then produces heatmaps highlighting the most influential image regions, enhancing transparency and trust in the predictions of the quantum model.

Example References

[1] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, “Quan- tum machine learning,” Nature, vol. 549, no. 7671, pp. 195–202, 2017.

[2] Q. Yang, Y. Liu, T. Chen, and Y. Tong, “Federated machine learning: Concept and applications,” ACM Transactions on Intelligent Systems and Technology, vol. 10, no. 2, 2019.