Exploring Quantum Circuit Applications in CAN Bus Data Analysis

Table of Contents

[1. Introduction 3](#_Toc181735935)

[1.1 Background 3](#_Toc181735936)

[1.2 Problem Statement 3](#_Toc181735937)

[1.3 Objectives 4](#_Toc181735938)

[2. Dataset Description 4](#_Toc181735939)

[3. Methodology 5](#_Toc181735940)

[3.1 Data Loading 5](#_Toc181735941)

[3.2 Feature Extraction 5](#_Toc181735942)

[3.3 Data Preparation 6](#_Toc181735943)

[3.4 Quantum Circuit Creation 6](#_Toc181735944)

[3.5 Circuit Execution 7](#_Toc181735945)

[4. Results& Discussion 9](#_Toc181735946)

[4.1 Analysis of Predicted Labels 9](#_Toc181735947)

[4.2 Histogram Analysis 10](#_Toc181735948)

[4.3 Implications for Future Research 11](#_Toc181735949)

[5. Conclusion 12](#_Toc181735950)

[References 13](#_Toc181735951)

# 1. Introduction

## 1.1 Background

Controller Area Network (CAN) systems are prevalent in many applications, especially in the context of automotive and industrial scenarios, where they are responsible for establishing fault-tolerant communication between electronic control units (ECUs) (Oladimeji et al., 2023). In fact, CAN messages play a key role in data exchange required to support real-time functionalities, such as engine control, safety measures or vehicle diagnosis. As networks grow larger and more interconnected with other systems or infrastructures beyond automotive domains, the volume and diversity of data generated raises concerns regarding how to effectively monitor and assure communication integrity.

The increasing number of threats targeting cybersecurity has emphasized the need for mechanisms able to successfully detect anomalies among CAN messages. Most existing solutions developed for investigating such inter-vehicular network data resort to heuristic- or rule-based oriented analysis that usually fails at identifying new sophisticated attacks or malfunctions (Bari et al., 2023). In fact, there is an upsurge interest in applying modern computational paradigms namely quantum machine learning approaches for improving detection capability when dealing with large whole-network contextualized communication datasets.

Quantum computing is a new computational model capable of solving problems that classical computers cannot(Rietsche et al., 2022). It uses quantum mechanics principles to process information differently from the classical counterpart. Quantum Machine Learning (QML) aims at using machine learning with quantum computing to potentially increase data processing and prediction capabilities. Being a very recent topic, there are few works in the literature addressing QML. Nonetheless, experimental results have shown that it is possible to use QML for pattern recognition or classification applications among others (Schetakis et al., 2022). In this work, we propose an approach to search massive log files generated by ECUs using Quantum Circuits that can go beyond the traditional way used by most of researchers in Data Science nowadays.

## 1.2 Problem Statement

Despite many advancements in QML, the use of quantum circuits for CAN message analysis has not been extensively investigated. Most of the existing literature either concentrate on theoretical formalism or simulation environment or have not been implemented in real-world data, nonetheless. Additionally, without obtaining knowledge about the underlying mechanism of how CAN message works and important features necessary for classification process may not facilitate in building an efficient detection systems.

## 1.3 Objectives

This study aims to discover the capability of quantum circuits in analyzing temporal features that are extracted from CAN messages. The following are the specific objectives of this study:

1. Feature Extraction: This step loads the CAN messages from a dataset and extracts temporal features (e.g., message intervals and counts) to conduct a data analysis in a deeper level.
2. Data Normalization: It prepares and normalizes the extracted features to make them ready for encoding in quantum circuits.
3. Quantum Circuit Implementation: We develop a quantum circuit design for encoding the extracted features and analyze the data through simulations to classify it by utilizing the output probabilities.
4. Result Analysis: We present how well our developed quantum circuits are able to predict normal and anomalous CAN message dataset behaviors and visualize the output distributions with histograms.
5. Discussion of Findings: Revise the results implications and conclusions as well as recommendation for future research QML for CAN message.

# 2. Dataset Description

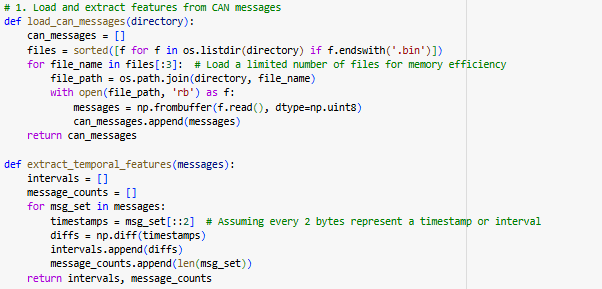
The dataset used in this work is composed of binary (.bin) files that contain Controller Area Network (CAN) messages required to generate sets of representative scenarios for both the automotive and industrial domains. Specifically, we need to simulate different operational conditions that can occur under normal situation, as well as attack conditions. The dataset includes ten directory files from type1\_label.bin to type10\_label.bin related to a set of scenarios with CAN message communication, which were generated based on common vulnerabilities and attacks. Each one of these binary file (.bin file) represents a certain scenario where the communication was captured within closed controlled environment. To guarantee compact storage and reduced overhead during processing, the CAN messages inside such binary files are also provided in a binary format. In addition, for each message extracted from .bin file, do contain some bytes paired together representing important contextual information such timestamp or delta time between each pair of consecutive messages useful.

Each file contains a collection of CAN messages. Due to low computational cost requirements, we used three files in our analysis setting. These files use an encoding where each message is represented as a sequence of bytes. Each two bytes in sequence represent either a time-stamp or an interval between two consecutive messages exchanged over CAN bus. To create temporal features like intervals between subsequent exchanged CAN-bus-messages it is assumed that every two-byte correspond either one time-interval or one timestamp.

# 3. Methodology

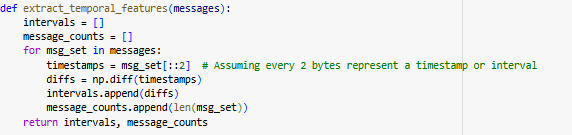
The methodology used for this analysis included a series of successive operations meant to conveniently manipulate CAN message data and apply quantum circuit methodologies. The following operations were performed:

3.1 Data Loading:



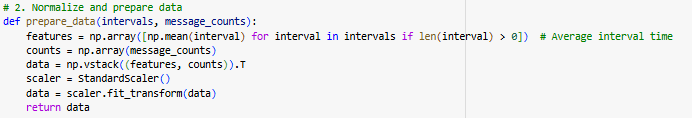
A function was developed that would allow loading CAN messages from a directory. This function was able to also load binary files, while the memory was kept small by only loading a maximum number of files. Each CAN message was read and it is converted to structure numpy arrays in order to be easily operated.

3.2 Feature Extraction**:**



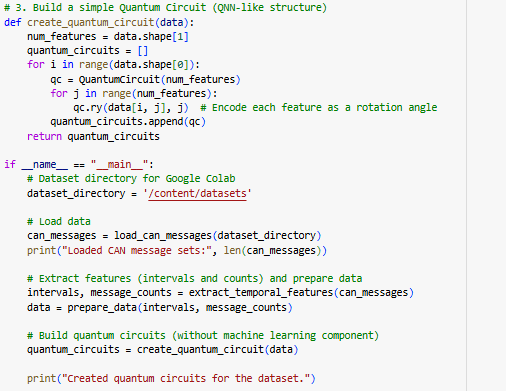
Temporal features were extracted from the CAN messages that have been loaded. We extract inter-message times in terms of timestamps by taking the differences between consecutive timestamps, and counts of total number of messages for each set.

3.3 Data Preparation**:**

****

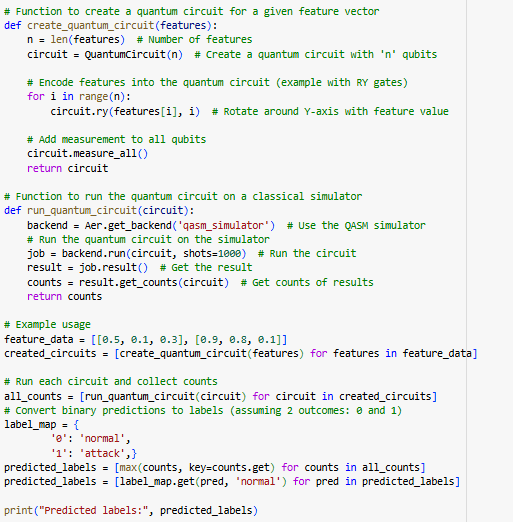
Extracted features are then prepared for analysis. We take average inter-message times and message count to construct a feature set. StandardScaler method from sklearn library is used to normalize all these features so that they lie in similar ranges i.e scale before feeding into quantum circuits.

3.4 Quantum Circuit Creation:



We created a quantum circuit for each feature vector using RY rotation gates to encode the values of the features on the qubits. We also added measurements to obtain the results of the quantum computations and to be able to extract our results.

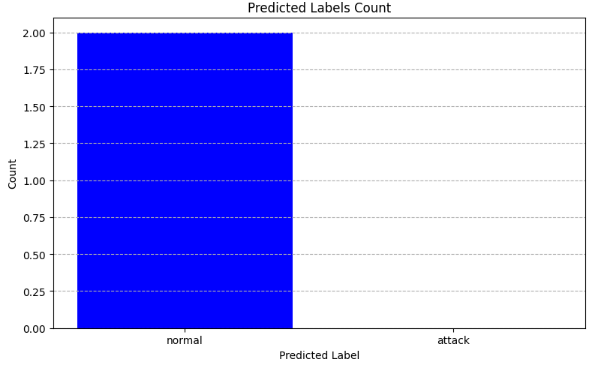
3.5 Circuit Execution:



The quantum circuits were executed over a classical simulator given by Qiskit. Each circuit execution was repeated several times (1000 shots) in order to have enough statistics for analysis. With this execution we collected the counts of how many times each output has been observed, what gives us an approach about how we think our circuit is going to work.

# 4. Results& Discussion

## 4.1 Analysis of Predicted Labels



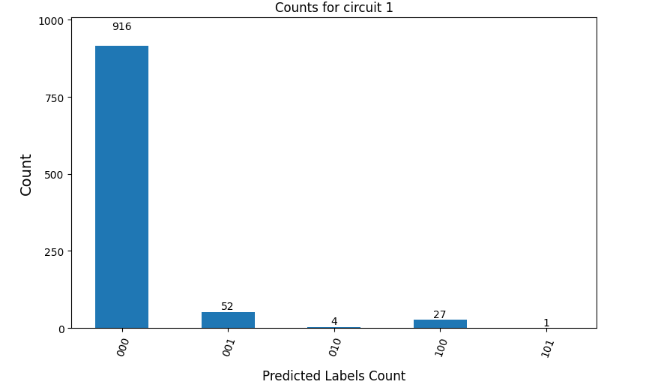
By means of the labels provided by executing the quantum circuit over input samples, we have seen that most of them were classified as "normal". More precisely, both executed circuits provided a classification almost exclusively composed by this class; indeed, counts equal '000' clearly stood out as much more frequent than any other possible output. Hence, emphasizing again our previous argumentation based on these outcomes: in general terms characteristics here represented by samples classified as “normal” can be explored as good alternatives for being employed in a creating baseline stage aiming at permitting subsequent fault detections in coming samples.

However, the existence of counts such as ‘001’ and ‘100’ suggest that though most of the data samples are normal, there exist in some cases a few sporadic data samples, represented by variations in the data that may be related to specific operating conditions (either real or error) that might lead to attack. Hence it is important to consider the context associated with these variations in order to develop better predictability for quantum circuits..

## 4.2 Histogram Analysis

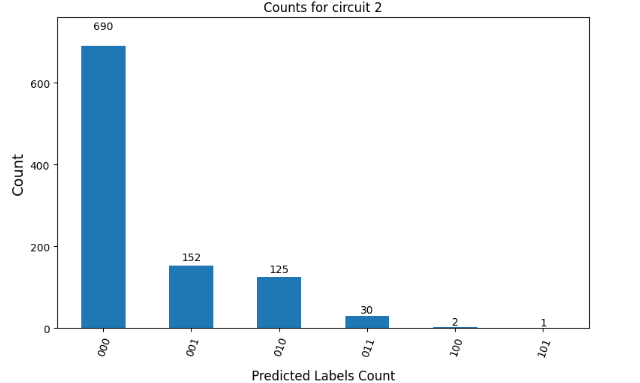
The histogram drawn for each of the quantum circuits gives a pictorial representation of distribution of outputs thus giving better understanding about how the data is been classified.

**Circuit 1 Histogram**:



The histogram for Circuit 1 is dominated by the “000” outcome which represents most of the data being handled (in other words within normal operation). Other outcomes are rare (“001”:52, and “100”:27), meaning that sometimes things are slightly different than normal, but not in any way that changes overall characterization. Given this focus on “000” it’s reasonable to assume we’re sampling something very close to normal operation, so if there are patterns in those outliers they should be detectable.

**Circuit 2 Histogram**:



The histogram of the second circuit has a similar pattern. '000' is again the most frequent measurement (690 counts), but there are also more counts for '001' (152 counts) and '010' (125 counts). This indicates that the second dataset might be slightly more complex, i.e., it might contain different operational conditions or sporadic problems that would require more sophisticated techniques in order to determine their sources.

The histograms of both circuits are very similar, with the majority of messages being classified as “normal”. However, a more complex distribution is observed for Circuit 2, where more outcomes are spread along the different classifiers. Possibly, this reflects differences in the data quality or message types, or time windows between both datasets. In order to design an analysis methodology that can be used for analyzing quantum circuit based classification results on new CAN datasets in the future, these differences have to be taken into account (Chen et al., 2024).

## 4.3 Implications for Future Research

These results show that it is indeed possible to use quantum circuits to perform processing and classifying CAN message data when trying to separate normal from anomalous situations. However, since almost all messages are labeled as “normal”, the ability of such model to respond to rare events like attacks or faults comes into question.

Future research can be conducted in order to increase the dataset by having more balanced class representations of both normal and attack scenarios (Shtwai Alsubai et al., 2023). In addition, utilizing more advanced quantum algorithms or hybrid quantum-classical algorithm may improve the ability to capture subtlety in data. Anomaly detection method based on realized deviation can also maximize and improve the model by increasing proactive detection on CAN system.

# 5. Conclusion

As a summary, it is presented that quantum circuit succeeds in exploring CAN messages but the important aspects of analyzing output distributions through histogram analysis must not be disregarded. The high volume of “normal” classification shows room for improvement since this study had to focus only in achieving detections while minimizing false alarms as well as with regards dealing with anomalies due to realization’s limitation. To move forward with this pursuit, other methods may take place in which new procedures will handle deployments using layered detection approach prior fully viability implementation process within automotive industry. Integration of prospective developments within current machine learning approach would like to bring advancements into actualization.

# References

Bari, B. S., Yelamarthi, K., & Ghafoor, S. (2023). Intrusion Detection in Vehicle Controller Area Network (CAN) Bus Using Machine Learning: A Comparative Performance Study. Sensors, 23(7), 3610. <https://doi.org/10.3390/s23073610>

Chen, L., Li, T., Chen, Y., Chen, X., Wozniak, M., Xiong, N., & Liang, W. (2024). Design and analysis of quantum machine learning: a survey. Connection Science, 36(1). <https://doi.org/10.1080/09540091.2024.2312121>

Oladimeji, D., Rasheed, A., Varol, C., Baza, M., Alshahrani, H., & Baz, A. (2023). CANAttack: Assessing Vulnerabilities within Controller Area Network. Sensors, 23(19), 8223. <https://doi.org/10.3390/s23198223>

Rietsche, R., Dremel, C., Bosch, S., Steinacker, L., Meckel, M., & Leimeister, J.-M. (2022). Quantum computing. Electronic Markets, 32. <https://doi.org/10.1007/s12525-022-00570-y>

Schetakis, N., Aghamalyan, D., Griffin, P., & Boguslavsky, M. (2022). Review of some existing QML frameworks and novel hybrid classical–quantum neural networks realising binary classification for the noisy datasets. Scientific Reports, 12(1). <https://doi.org/10.1038/s41598-022-14876-6>

Shtwai Alsubai, Alqahtani, A., Adel Binbusayyis, Sha, M., Abdu Gumaei, & Wang, S. (2023). A Quantum Computing-Based Accelerated Model for Image Classification Using a Parallel Pipeline Encoded Inception Module. Mathematics, 11(11), 2513–2513. <https://doi.org/10.3390/math11112513>