Pulsar anomaly detection: QOCSVM with CAE extracted features

Donovan Slabbert[1,*] and Francesco Petruccione[1,2]

[1]*School of Data Science and Computational Thinking and Department of Physics, Stellenbosch University, Stellenbosch 7602, South Africa*

[2]*National Institute for Theoretical and Computational Sciences (NITheCS), South Africa*

(Dated: May 24, 2024)

The hybrid field of quantum machine learning (QML) shows promise to improve upon classical machine learning in many ways; however, it still faces many limiting factors. Two of the biggest challenges facing quantum computing in the current NISQ era are noise and time. Noise, in the form of decoherence or qubit relaxation, as well as the time required for feature embedding, gate operations, and measurements, are well-known examples. This is accentuated by an increase in data size. As the data size increases, so does the number of qubits required for encoding. Larger quantum circuits with more qubits are more susceptible to noise and take longer to execute. These limitations significantly hinder the application of QML to larger datasets, specifically image datasets such as the HTRU-1 pulsar dataset, as embedding the individual pixels onto a quantum computer necessitates larger circuits.

Despite amplitude encoding partially mitigating these limitations by addressing the dimensionality problem through efficient data representation as amplitudes in a quantum state, dimensionality reduction may still be required for QML to be practical for large images and datasets. The dimensionality reduction should be applied without a significant loss of information. We rely on a feature extraction method that involves training a classical autoencoder for image reconstruction. This will serve as both a dimensionality reduction and feature extractor. Once the autoencoder successfully reconstructs the images, we isolate and flatten the latent space to serve as the new, reduced dataset. This latent space can then be used for further machine learning tasks, such as anomaly detection. Anomaly detection is a common machine learning problem where the goal is to identify a sample as a possible anomaly.

Combining this approach with the convenience that quantum support vector machines (QSVMs) typically require fewer training samples to learn patterns than variational approaches, we create a hybrid pipeline that integrates classical feature extraction using a trained convolutional autoencoder (CAE) for image reconstruction with a quantum-enhanced one-class support vector machine (QOCSVM) for anomaly detection. This method is applied to a real-world task of pulsar anomaly detection, using both the HTRU-1 image dataset and the 8-feature HTRU-2 dataset. Classical benchmarks and a supervised QSVM are also applied to the feature extracted data. Preliminary results show that the QOCSVM and QSVM perform comparably with its classical counterparts when meaningful features are extracted through image reconstruction. The hope is that this idea can be extended to larger datasets with more complicated images in the future.

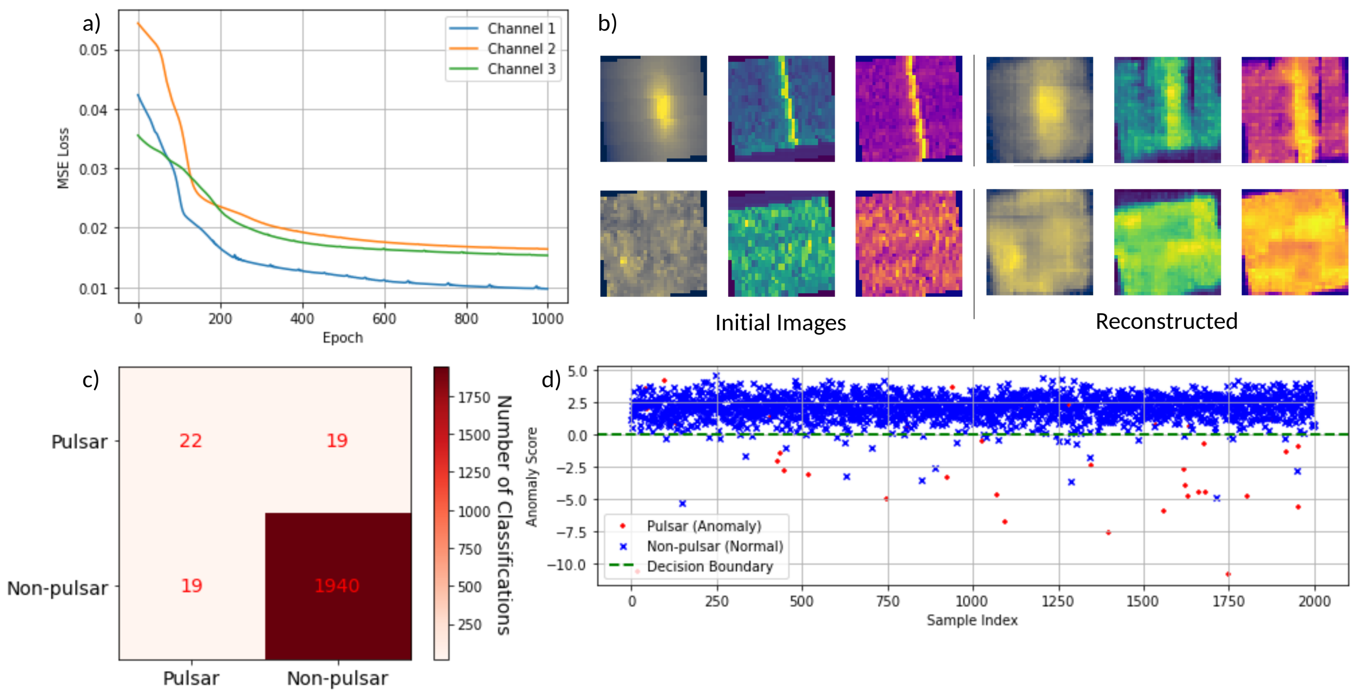

FIG. 1: Prelimminary results: a) Loss optimization and convergence for three classical autoencoders. b) Visualization of image reconstruction for both classes - anonaly and normal. c) Confusion matrix for a 2000 sample run. d) Visual representation of the separation between flagged anomalies and normal data.

*Electronic address: donovanslab@mweb.co.za