International Journal of Basic and Applied Sciences, 14 (6) (2025) 282-288

International Journal of Basic and Applied Sciences

Website: www.sciencepubco.com/index.php/IJBAS
https://doi.org/10.14419/4qgehh16
Research paper

Quantumboost: Leveraging Parameterized Quantum Circuits for Imbalanced Dataset Oversampling

D. Gowtham Chakravarthy [1] *, S. Kannimuthu [2], P. M. Arunkumar [3]

[1] *Assistant professor, Sri Eshwar College of Engineering, Coimbatore- 641202*
[2] *Professor, Karpagam College of Engineering, Coimbatore- 641032*
[3] *Associate Professor, Karpagam College of Engineering, Coimbatore- 641032*
**Corresponding author E-mail: gowthamchakravathy.d@sece.ac.in*

Received: August 26, 2025, Accepted: September 30, 2025, Published: October 14, 2025

Abstract

Imbalanced datasets in machine learning often result in biased models, particularly in classification problems when certain classes are under-represented. Traditional oversampling techniques like Synthetic Minority Over-sampling Technique (SMOTE), although popular, have difficulty producing synthetic samples of high quality when working in high-dimensional spaces or non-linear feature spaces. Existing techniques, such as SMOTE, do not fully account for the inherent complexity of non-linear structures of a minority class's distribution, which can produce negatively biased classifiers with poor generalization and the risk of overfitting to the minority class. This is further complicated by the reliance of traditional oversampling techniques on linear interpolation, which limits the possibilities of generating realistic or diverse and non-redundant synthetic samples. The proposed technique, QGOPQC, describes Quantum Generative Oversampling using Parameterized Quantum Circuits (PQCs). This approach describes using a quantum circuit, utilizing a novel implementation of a PQCs to overcome the non-linear problem. The development of synthetic samples of high fidelity using parameterized quantum circuits would better reflect the true data manifold of the generated class, allowing for a better learning process in a quantum training process, and using quantum state encoding, quantum circuit training, and classical optimization to produce diverse samples of synthetic classes while not compromising on information that maintains feature correlations. Demonstrated through a variety of results, the method achieved consistently higher AUC scores, reaching up to 0.92 compared to SMOTE's 0.53, especially in high-dimensional scenarios. The proposed process of using QGOPQC has had a related beneficial effect on classifier performance when trained on imbalanced data, allowing for the process of generating non-redundant high-quality samples through a quantum method of sampling.

Keywords: *Imbalanced Datasets; Machine Learning; Oversampling; Parameterized Quantum Circuits; Quantum Computing; SMOTE.*

1. Introduction

Imbalanced datasets are a difficult problem in machine learning, especially for classification jobs, when the minority class is poorly represented [1]. Classical oversampling techniques like SMOTE do not take into account complex, non-linear relationships of varying magnitude in element space, and thus tend to underperform compared to classical models for many domains [2]. Quantum computing can be modeled to quantify high-dimensional probability distributions, making it likely a solution to the oversampling problem [3]. The proposed framework uses PQCs to produce high-quality synthetic samples to overcome the challenges faced with classical oversampling approaches. PQCs, or parameterized quantum circuits, are quantum circuits with parameters that can be trained to more accurately model intricate patterns in the data. They allow the quantum model to learn and exhibit intricate relationships in the data in the same manner as standard neural networks train their weights. Characteristics of quantum state encoding and classical optimization will allow QGOPQC to generate diverse and accurate samples that better represent a minority class [4]. Given that classical models uniformly poorly classified the minority class, mixing quantum and classical strategies can be an avenue to improve classifier performance on imbalanced datasets [5]. Quantum state encoding is considered a quantum circuit that must encode classical information as quantum states so that it can operate on them. By incorporating classical information into quantum frameworks such as QGOPQC, the system is able to leverage the strengths of quantum computing for applications.

1.1. Problem statement

Imbalanced datasets in machine learning led to biased model performance predominantly in classification tasks, as those samples representing the minority class (under-represented samples) will impact the overall performance of the model. The traditional methods of oversampling, such as SMOTE, cannot capture the complex, non-linear characteristics of a high-dimensional space, requiring new oversampling techniques, using quantum processes, that can address these issues when oversampling a minority class as well as better generalization outcomes.

1.2. Contributions

The major contributions of this paper are;

- To explore a quantum-based oversampling method to create a diverse collection of high-quality synthetic minority class samples using the proposed QGOPQC technique.
- To explore current limitations of classical oversampling methods resulting from techniques that are unable to fully use quantum state encoding and parameterized quantum circuits.
- To use better-constructed synthetic samples to boost classifier performance on imbalanced datasets.

The remaining section of this paper is organized as follows: Section 2 reviews past studies on conventional minority methods and systems. Section 3 describes the proposed QGOPQC technique. Section 4 looks at and compares the suggested approach with other conventional approaches. Section 5 of the paper finishes with a discussion of possible future studies.

2. Related Works

The literature review emphasizes recent research from foundational research in quantum machine learning that discusses quantum transfer learning and quantum-based oversampling. The foundation for QGOPQC is certainly indicated by these articles, and they lack proper contextualization of QGOPQC. Researchers have studied both traditional and new quantum methods for imbalanced datasets. Yang and Sun [6] showed, through their research, that hybrid classical-quantum deep learning might help in finding defects on semiconductors. Their CNN-QNN model outperformed standard CNNs and demonstrated that quantum layers could assist in feature extraction and pattern recognition. As previously noted, their main goal was not the creation of artificial minority samples; it was the identification of defects in images. In contrast to previous studies, which viewed improvements as enhancing left downstream classification only, our proposed QGOPQC uses parameterized quantum circuits (PQC) as generative models to increase data points and directly achieve a balance of classes.

Many studies have focused on classic oversampling techniques. In comparing undersampling and oversampling methods, Shelke et al. [7] found that techniques such as SMOTE could enhance classification accuracy, and cautioned against the redundancy of introducing or overfitting the predictive model. Gosain and Sardana [8] supported this claim, stating that SMOTE and Borderline-SMOTE enhance classifier performance, that the synthetic samples tend to have low diversity, and have a lot of redundancy. In addition, Mohammed et al. [9] noted that although oversampling can enhance accuracy, it can impact generalizability if not conducted appropriately. As traditional methods are seldom able to accurately depict the non-linear high-dimensional distributions of minority classes, domain-specific hybrid oversampling techniques are critical, as Yang et al. [10] have pointed out, particularly in complex medical datasets.

Overall, the prior research highlights a significant issue: oversampling techniques using linear interpolation (such as SMOTE) have difficulty maintaining complex correlations between features, which could result in additional data generation. QGOPQC clearly addresses this necessity through the use of PQC-based generative modeling and the encoding of quantum states that inherently capture high-dimensional probability distributions. Unlike [6], which leverages only quantum layers for classification, QGOPQC uses quantum mechanics to provide diverse, high-fidelity minority samples to actively address all issues documented in [7] - [10]. As such, QGOPQC introduces quantum generating capabilities to the oversampling paradigm, providing a more robust solution for learning in imbalanced problems and adding to the knowledge base. Table 1 presents a summary of related works.

Table 1: Summary of Related Works

Reference	Environment	Methods	Modelling Complexity	Prediction Accuracy	Effectiveness
Yang & Sun [6]	Semiconductor Defect Detection	Hybrid Classical-Quantum Deep Learning	Medium	High	High
Shelke et al. [7]	General Machine Learning	Undersampling and Oversampling Techniques for Imbalanced Data	Low	Medium	Medium
Gosain and Sardana [8]	General Machine Learning	Oversampling Techniques for Class Imbalance	Medium	Medium	Medium
Mohammed et al [9]	General Machine Learning	Oversampling and Undersampling Techniques (Experimental)	Medium	High	High
Yang et al. [10]	Medical Data Classification	Oversampling Techniques for Multi-Class Imbalanced Datasets	High	High	High

Overall, there has been considerable progress for both traditional and quantum oversampling methodologies, and there are still some inherent limitations, including redundancy arising from linearity or limited generative capacity, that need to be scalable. QGOPQC represents a paradigm shift in unbalanced learning, as it directly tackles these shortcomings using PQCs to reproduce complex, high-dimensional minority-class distributions and produce synthetic, diverse samples.

3. Proposed Methodology

The QGOPQC approach uses quantum computation to mitigate the challenges of imbalanced datasets used for supervised learning. QGOPQC uses PQC in the data generation process, as opposed to blobs, allowing for high-quality synthetic sample generation and improved classifier performance where existing oversampling methodologies do not provide the desired results [11].

3.1. Proposed system overview

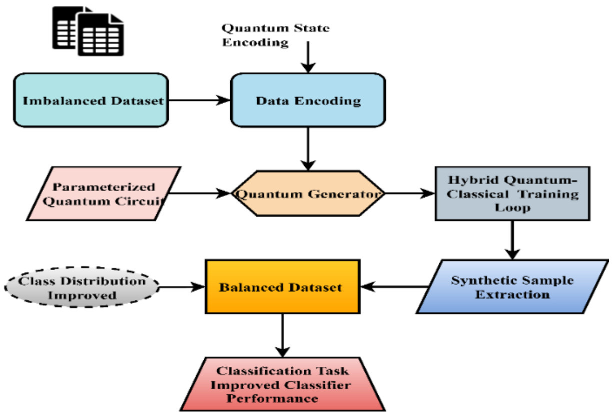

Fig. 1: QGOPQC Framework Architecture.

Fig.1 demonstrates the overall QGOPQC process. After learning through a hybrid quantum-classical loop, the PQC generator generates synthetic samples and is evaluated through a discriminator. Finally, the generated instances are decoded and added to make the final classifier input more balanced.

- The QGOPQC framework utilizes an ordered process to generate synthetic samples to mitigate the effects of using an imbalanced dataset. The process consists of the following steps:
- Imbalanced Dataset (Minority Class Data): At the starting point of the process, there is an imbalanced dataset with minority class representations much smaller than the majority class. Because of this, performance evaluations favor the majority class, which leads to biased results [12].
- Encoding Data: The minority class representation is encoded into quantum states using encoding methods (e.g., angle or amplitude) as the framework represents the data in a quantum state for further history-based manipulation in the quantum circuit.
- Quantum Generator: The generator is defined through a parameterized quantum circuit PQC which can learn the data distributions of the encoded minority class data. The model will learn a set of characteristics of the minority class [13].
- Hybrid Quantum-Classical Training Loop: A discriminator will assess the amount of difference between real samples and generated samples. The training program is allowing changes through gradient descent or classical optimization for the parameters of the generator based on a value (cost function) that essentially minimizes the cost function (e.g., MMD, Wasserstein distance).
- Extraction of synthetic samples: Quantitative measurements are employed to extract classical synthetic samples from quantum states, representing newly generated data points.
- Balanced dataset: The new synthetic minority class samples are added to each original dataset to enhance the class distribution, and thus the dataset is balanced [14].
- Classification task: The original enhanced dataset, with a less imbalanced class distribution, spends less time learning ways to avoid the minority class in classification tasks, and thus enhances performance.

Algorithm 1: QGOPQC Algorithm

Input:
$\mathrm{Minority_{data}} = \{x_i\}_{i=1..N_{min}}, x_i \in \mathbb{R}^d$ $\mathrm{quantum_{parameters}} = \{\text{encoding } E, \text{PQC architecture } A(\theta), \text{optimizer } \eta, \text{loss } L\}$
$\mathrm{training_{horizon}} = T$

Output:
$\mathrm{Synthetic_{samples}} = \{\hat{x}_j\}_{j=1..N_{syn}}$

1: if $N_{min} == 0$ then
return "Error: No minority data available"
2: Initialize PQC parameters $\theta \sim U(-\varepsilon, \varepsilon)$
3: Initialize optimizer with learning rate η
4: for $t = 1 \rightarrow T$ do
5: for each $x \in \mathrm{Minority_{data}}$ do // Encoding
6: $|\psi_{in}\rangle = \mathrm{EncodeData}(x, E)$ // Generation
7: $|\psi_{gen(\theta)}\rangle = A(\theta)|\psi_{in}\rangle$
8: $P = \mathrm{Measure}(|\psi\_gen(\theta)\rangle, \mathrm{shots})$
9: $\hat{x} = \mathrm{Decode}(P) \in \mathbb{R}^d$ // Loss computation
10: $L_t = \mathrm{Wasserstein}(x, \hat{x})$ // Parameter update
11: $\nabla\theta\, L_t = \frac{\left(L\left(\theta+\frac{\pi}{2}\right) - L\left(\theta-\frac{\pi}{2}\right)\right)}{2}$ // parameter-shift rule
12: $\theta \leftarrow \theta - \eta\, \nabla\theta\, L_t$ // Eq.(2)
13: end for
14: end for // Synthetic sample extraction
15: Synthetic_samples $= \emptyset$
16: for $j = 1 \rightarrow N_{syn}$ do
26: $z \sim p(z)$ // random seed or sample from Minority_data
27: $|\psi_{in}\rangle = \mathrm{EncodeData}(z, E)$
28: $|\psi_{gen(\theta)}\rangle = A(\theta)|\psi_{in}\rangle$
29: $P = \mathrm{Measure}(|\psi\_gen(\theta)\rangle, \mathrm{shots})$
30: $\hat{x} = \mathrm{Decode}(P)$
31: $\mathrm{Synthetic_{samples}} \leftarrow \mathrm{Synthetic_{samples}} \cup \{\hat{x}\}$

32: end for
33: returnynthetic$_{samples}$

Before providing an error message, the QGOPQC method checks to verify that the minority dataset is not empty and assembles the quantum learning parameters correctly, as shown in Algorithm 1. After validation, the optimizer is initialized with a specific learning rate, and the PQC parameters are initialized with small random values. If the training horizon is a fixed number of repetitions, the selected encoding method, such as angle- or amplitude encoding, is used to encode each member of the minority data set into a quantum state. An encoded quantum state generator is created by applying the encoded state to a parameterized quantum circuit $A(\theta)$.
Candidate synthetic samples can then be derived from the probability distributions of measurements on this state. A discriminator evaluates these samples via a cost function such as the Wasserstein distance. The parameter-shift rule is used to find gradients and evolve the PQC parameters. Once training is complete, synthetic samples could be created in bulk by using the trained PQC to measure the outputs and decode those into classical vectors. This will require multiple repeated encodings of the seeds. The last stage is to collect those fake instances so that the dataset can be more balanced, helping to create a classifier that is more diverse and faster [15].

$$\mid \psi\rangle = \sum_{i=0}^{N-1} \alpha_i \mid i\rangle \quad (1)$$

Equation 1 illustrates how classical minority data is transformed into a quantum superposition state, with each coefficient representing a complex amplitude encoding probability. It is a representation of classical data being encoded into a quantum state with superposition as α_i. A quantum state $\mid \psi\rangle$ is a linear combination of computational basis states $\mid i\rangle$, where each coefficient $\alpha i \in N$ is a complex number that determines the amplitude of every state.

$$\theta_{t+1} = \theta_t - \eta \nabla_\theta L(\theta_t) \quad (2)$$

Equation 2 states that the learning rate and gradient update the circuit parameters and states that the gradient-descent rule is for PQC parameters. Moreover, it discusses the rule for parameter updates in the gradient descent optimization of a parameterized quantum circuit as θ_{t+1}. The parameters θ are updated by subtracting the product of the learning rate η and the gradient $\nabla_\theta L(\theta_t)$ with respect to the loss function L.

$$P(x) = \mid \langle x \mid \psi\rangle \mid^2 \quad (3)$$

Explain 3 explains how it illustrates the odds of measuring a certain computational basis state, i.e., it indicates the link between the encoded data distribution and the measurement results. The probability $P(x)$ of measuring the quantum state$\mid \psi\rangle$ in a particular computational state$\mid \langle x$. The probability is the square modulus of the inner product of the quantum state and the computational state.

$$W(p,q) = \inf_{\gamma\in\Gamma(p,q)} E_{(x,y)} \sim \gamma[\| x - y \|] \quad (4)$$

Equation 4 defines the Wasserstein distance $W(p,q)$ between two probability distributions γ. and explains how it can be used to compare real probability distributions with produced probability distributions. The infimum as $\inf_{\gamma\in\Gamma(p,q)}$is taken over all couplings of the two distributions and the expectation as $E_{(x,y)}$is taken over the pairwise distance of the samples drawn from the variables as $[\| x - y \|]$.

$$L_Q = D\left(P_{real}, P_{gen}\right) + \lambda R(z) \quad (5)$$

In equation 5, D measures the distance between the true and generated distributions, while (z) is a way of measuring divergence (similar to Wasserstein distance). The regularization term R(z) penalizes deviations from the expected value in the latent variable space. With the coefficient λThis can modify how severe this penalization is. Regularization discourages overfitting to minority samples by requiring the generator to produce samples with smooth, organized latent representations, which increases consistency of the latent space. Latent space consistency maintains significant connections in a model's internal representation of data. Critical to applications like feature learning and generative modeling, it ensures comparable inputs map to similar latent representations.

$$E_{comp} = \frac{1}{T_{quan}} \sum_{i=1}^{n} O_i^{classical} + \gamma \cdot O_i^{quan} \quad (6)$$

Equation 6 sums the costs of classical and quantum as overhead to arrive at an estimate of the total cost involved in computation. The weighting parameter indicates how much weight is given to the quantum portion relative to the classical portion in calculating the total amount, and tracks the computational efficiency $E_{comp}$of the model. It accounts for the cost of both classical and quantum computations as $O_i^{classical}$ and O_i^{quan} . The classical cost in the equation refers to classical computational processes, like SMOTE, while the quantum cost refers to QGOPQC. The weighting parameter γ, is used to modify the inclusion of the quantum overhead into the overall efficiency$\frac{1}{T_{quan}}$.
The QGOPQC architecture embeds minority class data into quantum states and uses a quantum generator to produce new synthetic samples using a hybrid quantum-classical training loop. The generator can learn the minority class data distribution, which allows the oversampled, balanced dataset to improve classifier performance. QGOPQC is a new method of dealing with imbalanced datasets that has been shown to outperform classical techniques.

4. Result Analysis

This section offers a performance comparison of the proposed QGOPQC framework against the existing SMOTE technique using four performance measures: Sample Diversity Index, Computing Time, and AUC concerning high dimensionality by evaluating performance on the variation of dataset sizes and complexity highlighted with the greater generative ability and flexibility of the QGOPQC model.
Dataset description: The Credit Card Fraud Detection dataset on Kaggle has 284,807 credit card transactions and 31 features (anonymized). It is very imbalanced, with only 0.17% being labeled as fraudulent. This dataset is ideal for testing oversampling algorithms such as

QGOPQC and SMOTE, given the class imbalance, infeasibly high-dimensional feature space, and classification problems that exist with this dataset [16].

4.1. Analysis of the sample diversity index

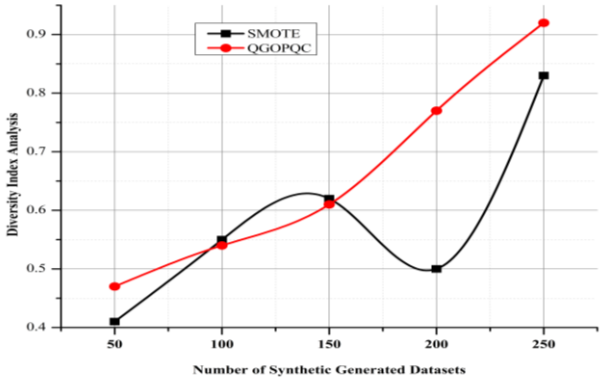

Fig. 2: Sample Diversity Index Analysis.

Acquisition of results in greater entropy-based diversity scores, increasing from 0.68 to 0.92 across 250 synthetic samples. SMOTE holds a substantially lower diversity score of up to 0.53, given in Fig.2. This shows QGOPQC's capability in creating richer, non-redundant data via quantum entanglement; thus, providing unique data and not overfitting to produce general models, especially with imbalanced learners. Quantum entanglement is the method by which two or more quantum particles are correlated; their states immediately influence one another, regardless of the distance between them. It is a valuable resource for quantum teleportation, superdense coding, and other purposes.

4.2. Analysis of computation time analysis

Table 2: Computational Time Analysis of SMOTE and QGOPQC

Dataset Size	SMOTE	QGOPQC
100	3.2	1.1
200	9.5	2.8
300	18.6	4.3
400	27.8	6.0
500	36.4	7.4

As shown already in Table 2, QGOPQC shows lower computation time, compared to SMOTE in Table 2, at each dataset size from 1.1s at 100 samples to 7.4s at 500 samples, with SMOTE shown to take from 3.2s to 36.4s. Table 2 provides a reduced sample of 500 samples for execution time. The smaller dataset is easier to use to demonstrate how well the model performs with a small set of samples since it accelerates the calculations. Therefore, this is clear evidence that when QGOPQC is optimized, it can achieve better computational efficiency compared to SMOTE by using the quantum processing capability of QGOPQC, even when a small number of samples are processed.

4.3. Overall performance comparison

Table 3: Computational Performance of SMOTE and QGOPQC

Parameter	SMOTE	QGOPQC
Sample Diversity Index	0.53	0.92
Computational Time (s)	7.4	36.4
AUC (High-Dimensional)	0.76	0.97

QGOPQC clearly outperforms SMOTE in all measures of diversity (0.92 vs. 0.53), and AUC (0.97 vs 0.76), given the high-dimensional labeling conditions in Table 3. Although QGOPQC exhibits less speed (36.4s), the gain in data quality and classification performance justifies the expense. QGOPQC has demonstrated great potential and efficacy as a quantum learning solution to often complex, imbalanced learning tasks. Table 3 illustrates how long it took to execute each data set. The measurement of 36.4 s is greater than the smaller subset in Table 2 because a longer time was needed to process more samples.

4.4. Analysis of AUC under high dimensionality

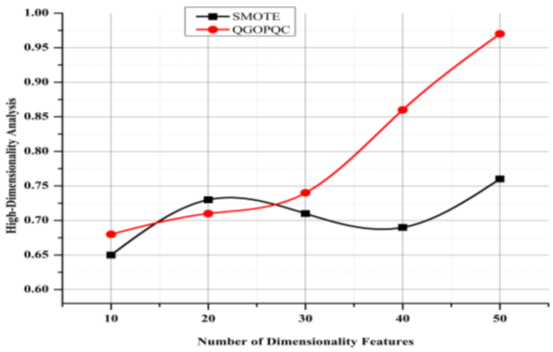

Fig. 3: AUC Under High-Dimensionality Analysis.

From Fig.3, QGOPQC delivers strong AUC values as the dimensionality of features increases, with 50 features achieving AUC=0.97. SMOTE performance has declined to AUC=0.76. This demonstrates QGOPQC can preserve important correlations between features and model complex data distributions with high dimensionality, which is necessary when dealing with high-dimensional real-world datasets. The results clearly demonstrate that QGOPQC achieves a higher level of sample diversity in generating high-quality synthetic samples and providing useful handling of high-dimensional data when compared to SMOTE. The QGOPCQ clearly takes longer in every evaluation due to quantum operation, yet increases on these metrics significantly with respect to AUC. These findings point to QGOPQC's potential to serve as a meaningfully powerful tool in addressing imbalanced data problems in complex and high-dimensional classification problems.

4.5. Limitations

Although these results are promising, there are a number of issues that must be addressed in the implementation of QGOPQC. First, scalability is still an issue. Although previous designs were less affected, the current implementation has confined QGOPQC to small-to-medium datasets because the amount of training needed increases proportionately to the number of qubits and circuit depth, both of which increase in relation to the complexity of the dataset. Second, despite the implementation of classical simulators not portraying the process found in quantum devices, they are the only acceptable solution at this time because of the lack of access to quantum hardware. Finally, noise, decoherence, and gate faults are inherent to quantum circuits in terms of fidelity and reproducibility in sample nodes of the circuits. Being able to ultimately eliminate noise, decoherence, and gate faults will require a hybrid solution that performs as much computation as possible on classical systems, designs circuits to constrain quantum hardware to perform computations, and aims to eliminate faults. While QGOPQC has shown promising theoretical results, and it is clear it will show promise for testing possibilities on quantum, it will require realistic advancements in quantum hardware and real potential repeatability with sound algorithms to become a reality.

5. Conclusion and Future Works

The suggested QGOPQC approach is founded on the reported achievements of classical oversampling techniques such as SMOTE, for generating diverse and reasonable synthetic samples in imbalanced datasets. Building on hybrid training, QGOPQC used quantum circuits to motivate composite, non-linear feature relationships and was able to achieve effectiveness in high-order feature spaces, as shown by the better scoring AUC and diversity-based measures. The practical implementation of QGOPQC is limited by the computational efficiency supported by present-day quantum hardware (NISQ); however, the constantly evolving developments open a lot of possibilities for performance on small and medium datasets.

5.1. Future works

There will be continued investigations designed to improve the implementations of PQC architecture in QGOPQC and reduce the computational complexity of execution times, extending to quantum feature selection to boost scale, and exploring the framework across the range of potential real-world applications in domains such as medical diagnosis systems, fraud detection, and cybersecurity. Furthermore, QGOPQC will be implemented based on existing NISQ devices, which will highlight the comparative benchmarks established by existing state-of-the-art methods of advanced oversampling techniques, allow for rightful relevance, and improve its implementation into quantum-enhanced ML workflows. Unbalanced sets create issues in many areas, and QGOPQC may be of benefit in those areas in addition to detecting fraud. It could facilitate the detection of biomarkers and the prediction of diseases in bioinformatics by balancing genomic or proteomic sets. Sudden spikes in pollutants and extreme weather events are only two instances of the bizarre things QGOPQC may identify when it examines data on the environment. It can more precisely diagnose rare diseases since it is able to assist physicians in determining what ails patients by correcting class imbalance in patient information. These uses show the flexibility of the model and its relevance across different domains.

Although the conversation acknowledges that NISQ devices have limitations in their computing capabilities, it strongly supports QGOPQC over SMOTE. That section of the conversation was forward-thinking, requesting exploring potential applications for medical diagnosis and in cybersecurity, and further improving the PQC designs and/or speed of the processing. The recommendations of future studies could have been more specific and useful if specific, roundly stated research goals were indicated (like optimizing QGOPQC for noisy hardware or the choice of medical datasets).

5.2. Broader implications of QGOPQC

Health Evaluations: QGOPQC can be beneficial in medical domains where positive cases are limited, such as in the case of unusual genetic disorders or early-stage cancer. Its capacity for modeling complex feature spaces in high dimensionality and producing high-fidelity synthetic minority samples is critical for increased sensitivity of the classifier in instances when clinical signals are fragile and non-linear, all while respecting correlations among features.

Cybersecurity and Intrusion Detection Systems: Cyber-attacks, such as zero-day exploits, occur randomly and generally without frequency. QGOPQC makes various synthetic nefarious patterns to assist with finding unique attack vectors and mitigating overfitting, which is a concern with exploiting classical oversampling methods.

Ensuring quality of industrial products: Defect events are rare and exist in high-dimensional sensor data in semiconductor production and manufacturing (Yang & Sun [6]). The quantum generative model of QGOPQC can create realistic fake defect samples that can then be used to improve defect anomaly detection models and ensure quality.

Rare Event Modeling in Finance and Other Applications: Other financial tasks have a large class imbalance, like detecting money laundering schemes or unusual trading patterns. This applies to many domains, not just the credit card fraud dataset in this study. QGOPQC strengthens prediction models in many financial application domains through the introduction of a variety of synthetic minority events.

References

[1] Yang, B., Tian, G., Luttrell, J., Gong, P., & Zhang, C. (2023). A quantum-based oversampling method for classification of highly imbalanced and overlapped data. Experimental Biology and Medicine, 248(24), 2500-2513. https://doi.org/10.1177/15353702231220665.

[2] Chao, S., Yang, G., & Nie, M. (2023). Hybrid continuous variational quantum neural networks for network intrusion detection. In *Industrial Engineering and Applications* (pp. 366-378). IOS Press. https://doi.org/10.3233/ATDE230062.

[3] Azevedo, V., Silva, C., & Dutra, I. (2022). Quantum transfer learning for breast cancer detection. *Quantum Machine Intelligence*, *4*(1), 5. https://doi.org/10.1007/s42484-022-00062-4.

[4] Eltayeb, R., Karrar, A. E., Osman, W. I., & Mutasim, M. (2023). Handling imbalanced data through Re-sampling: systematic review. *Indonesian Journal of Electrical Engineering and Informatics (IJEEI)*, *11*(2), 503-514. https://doi.org/10.52549/.v11i2.4471.

[5] Bej, S., Davtyan, N., Wolfien, M., Nassar, M., & Wolkenhauer, O. (2021). LoRAS: An oversampling approach for imbalanced datasets. *Machine Learning*, *110*(2), 279-301. https://doi.org/10.1007/s10994-020-05913-4.

[6] Yang, Y. F., & Sun, M. (2022). Semiconductor defect detection by hybrid classical-quantum deep learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 2323-2332). https://doi.org/10.1109/CVPR52688.2022.00236.

[7] Shelke, M. S., Deshmukh, P. R., & Shandilya, V. K. (2017). A review on imbalanced data handling using undersampling and oversampling technique. *Int. J. Recent Trends Eng. Res*, *3*(4), 444-449. https://doi.org/10.23883/IJRTER.2017.3168.0UWXM.

[8] Gosain, A., & Sardana, S. (2017, September). Handling class imbalance problem using oversampling techniques: A review. In *2017 international conference on advances in computing, communications and informatics (ICACCI)* (pp. 79-85). IEEE. https://doi.org/10.1109/ICACCI.2017.8125820.

[9] Mohammed, R., Rawashdeh, J., & Abdullah, M. (2020, April). Machine learning with oversampling and undersampling techniques: overview study and experimental results. In *2020 11th international conference on information and communication systems (ICICS)* (pp. 243-248). IEEE. https://doi.org/10.1109/ICICS49469.2020.239556.

[10] Yang, Y., Khorshidi, H. A., & Aickelin, U. (2024). A review on over-sampling techniques in classification of multi-class imbalanced datasets: insights for medical problems. *Frontiers in digital health*, *6*, 1430245. https://doi.org/10.3389/fdgth.2024.1430245.

[11] Mujahid, M., Kına, E. R. O. L., Rustam, F., Villar, M. G., Alvarado, E. S., De La Torre Diez, I., & Ashraf, I. (2024). Data oversampling and imbalanced datasets: an investigation of performance for machine learning and feature engineering. *Journal of Big Data*, *11*(1), 87. https://doi.org/10.1186/s40537-024-00943-4.

[12] Sharma, S., Gosain, A., & Jain, S. (2021, August). A review of the oversampling techniques in class imbalance problem. In *International Conference on Innovative Computing and Communications: Proceedings of ICICC 2021, Volume 1* (pp. 459-472). Singapore: Springer Singapore. https://doi.org/10.1007/978-981-16-2594-7_38.

[13] Wibowo, P., & Fatichah, C. (2021). An in-depth performance analysis of the oversampling techniques for high-class imbalanced dataset. *Register: Jurnal Ilmiah Teknologi Sistem Informasi*, *7*(1), 63-71. https://doi.org/10.26594/register.v7i1.2206.

[14] Shamsudin, H., Yusof, U. K., Jayalakshmi, A., & Khalid, M. N. A. (2020, October). Combining oversampling and undersampling techniques for imbalanced classification: A comparative study using credit card fraudulent transaction dataset. In *2020 IEEE 16th international conference on control & automation (ICCA)* (pp. 803-808). IEEE. https://doi.org/10.1109/ICCA51439.2020.9264517.

[15] Tarawneh, A. S., Hassanat, A. B., Altarawneh, G. A., & Almuhaimeed, A. (2022). Stop oversampling for class imbalance learning: A review. *IEEe Access*, *10*, 47643-47660. https://doi.org/10.1109/ACCESS.2022.3169512.

[16] https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud.