IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 15 August 2024, accepted 7 November 2024, date of publication 13 November 2024,
date of current version 25 November 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3496901

RESEARCH ARTICLE

Quantum Autoencoder for Enhanced Fraud Detection in Imbalanced Credit Card Dataset

CHANSREYNICH HUOT[1], SOVANMONYNUTH HENG[1], TAE-KYUNG KIM[2], AND YOUNGSUN HAN[1], (Member, IEEE)

[1]Department of AI Convergence, Pukyong National University, Nam-gu, Busan 48513, South Korea
[2]Department of Management Information Systems, Chungbuk National University, Cheongju-si, Chungcheongbuk-do 28644, South Korea

Corresponding author: Youngsun Han (youngsun@pknu.ac.kr)

This research was supported by ''Regional Innovation Strategy (RIS)'' through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (MOE) (2023RIS-007).

ABSTRACT Credit card fraud detection is crucial for financial security which entails identifying unauthorized transactions that can result in significant financial losses. Detection is inherently challenging due to the rarity and indistinguishability of fraudulent transactions from genuine ones, which makes it an anomaly detection problem. Traditional detection systems struggle with the highly imbalanced nature of transaction datasets, where genuine transactions vastly outnumber fraudulent cases. In response to these challenges, we propose a novel detection model utilizing Quantum AutoEncoders-based Fraud Detection (QAE-FD). Our approach leverages quantum computing principles to enhance anomaly detection capabilities by encoding transaction data into compressed quantum states and optimizing the model against a loss function that evaluates the fidelity in flagging fraudulent transactions. The efficacy of the QAE-FD model is tested on a real-world credit card transaction dataset, achieving a G-mean of 0.946 and an AUC of 0.947 which demonstrates superior performance compared to existing models. Our results indicate that QAE-FD has not only higher accuracy in fraud detection but also better computational efficiency. The integration of quantum autoencoders is a promising advancement in the field of anomaly detection for credit card fraud, addressing the limitations of imbalanced datasets and offering a scalable solution for real-time detection systems.

INDEX TERMS Anomaly detection, credit card fraud detection, imbalanced dataset, quantum autoencoder (QAE), quantum machine learning (QML).

I. INTRODUCTION

Credit card fraud is a growing problem with extensive financial implications for corporations, government bodies, and individuals. The primary method of this fraud is the exploitation of security vulnerabilities, especially through the misuse of stolen credit cards. This not only creates profound challenges in thwarting such fraudulent activities but also undermines public trust, potentially destabilizing economic systems and impacting the broader cost of living [1]. Given that fraud detection inherently falls within the realm of anomaly detection, the methodologies applied in this field are critical for effective fraud mitigation [2]. Machine learning algorithms are traditionally successful in anomaly detection within classical datasets. This success has prompted inquiries into the potential applicability of quantum machine learning (QML) algorithms to anomaly detection within quantum systems. Autoencoders (AEs), for instance, are increasingly used to identify anomalies directly due to their efficacy in feature compression and reconstruction [3]. The growing volume of transactions, often from uncertain or unverified sources, underscores the urgency for financial organizations to detect fraudulent activities robustly. Unlike canonical datasets, such as the Modified National Institute of Standards and Technology [4] and Iris [5] datasets, financial transaction datasets are characteristically unbalanced, posing unique challenges for traditional data analysis [6]. The emergence of QML has sparked interest due to its potential to expedite

The associate editor coordinating the review of this manuscript and approving it for publication was Yongming Li.

various data-driven tasks, including anomaly detection [7], [8]. Quantum autoencoders (QAEs) in particular have shown potential in managing high dimensional and outperforming classical approaches using the inherent efficiencies of quantum computation [9]. QAEs are effective in anomaly detection, which is crucial for fraud identification. Empirical studies across sectors, including financial fraud and medical anomalies (e.g., breast cancer detection), have obtained promising outcomes using QAEs [10].

Anomaly detection research has led to the development of robust methodologies using both classical and quantum approaches. Significant works in this area include "Quantum Machine Learning for Quantum Anomaly Detection," [7] which explores the capabilities of deep learning and quantum algorithms in identifying outlying data. Furthermore, QAEs have been used to detect anomalies in various scenarios. For instance, [11] proposed a QAE-based method for detecting anomalous phases in the context of quantum Hamiltonian problems. Then, [12] proposed the variational quantum one-class classifier, which simplifies the QAE structure by primarily utilizing its encoder component. This model outperformed a classical autoencoder (CAE) and was comparable with a quantum one-class support vector machine (QO-SVM) in most cases under similar training conditions. Moreover, a hybrid quantum model was proposed by addressing the quantum-classical methods for fraud detection that have been explored; this hybrid model combines classical models, such as random forests, with quantum support vector machines (SVMs) to harness quantum computing's potential for feature selection and enhance fraud detection accuracy [13]. These studies laid the groundwork for our investigation of the application of QAE to credit card fraud detection. We aim to harness these advanced computational techniques to address the critical challenge of identifying fraudulent transactions within highly unbalanced datasets. The choice of Quantum Autoencoders (QAEs) over classical methods is driven by their unique ability to leverage quantum parallelism, allowing for more efficient compression of high-dimensional data into lower-dimensional quantum states. Unlike classical autoencoders, which rely on traditional computational frameworks, QAEs exploit the inherent parallelism of quantum systems to process and encode information simultaneously across multiple states. This capability makes QAEs particularly effective in handling high-noise, high-dimensional datasets, where classical models often struggle with scalability and performance degradation. By encoding data into compact quantum representations, QAEs can enhance anomaly detection by isolating subtle deviations that might be obscured in noisy environments, ultimately leading to more accurate and robust detection of anomalies in complex datasets.

In this study, we propose a novel Quantum AutoEncoders-based Fraud Detection (QAE-FD). This framework is specifically designed to address the challenges posed by the highly unbalanced datasets (credit card fraud problem) prevalent in financial transaction data. At the core of our approach, we use a quantum circuit that functions as an encoder to compress transactional data into a quantum state of a lower dimension. This state is then used to identify fraud by assessing deviations from the expected quantum state of legitimate transactions. The novelty of our method lies in its dual-phase operational structure. Initially, the quantum encoder maps classical transaction data into a quantum Hilbert space. Then, it deploys a quantum classifier that evaluates the fidelity of the encoded state against predefined thresholds to detect potential fraud. This approach enhances detection sensitivity and significantly accelerates computation, leveraging the parallel processing capabilities of quantum computing. By integrating classical data preprocessing with quantum-based anomaly detection, our model offers a robust solution to the complexities of fraud detection in large-scale financial datasets. This method redefines the standards for accuracy and efficiency in the field, promising a significant advancement in the fight against credit card fraud. The contributions of this paper are summarized as follows:

- We propose a model for detecting fraudulent transactions that adapts the QAE model by using the quantum encoder part only.
- We evaluate our proposed model using a real, highly imbalanced financial (credit card) dataset with the adaption to real-world application scenarios.
- We perform sensitivity analysis on the hyperparameters of our proposed model, including circuit layers and the number of thresholds.
- We enhance the interpretability of our proposed model by demonstrating the necessary metrics, achieving a geometric mean (G-mean) of 0.946 and an area under the curve (AUC) of 0.974 in the most sensitive case, where the threshold is set to 0.7.

The remainder of this paper is organized as follows. In Section II, we explain the background. Section III about related existing studies. Section IV shows the details of our proposed method. In Section V, we present the experimental setup and results. Section VI covers the sensitivity analysis of the study. Finally, the conclusion of our study is in VII.

II. BACKGROUND

In this section, we explain the credit card fraud detection problem, classical, and quantum autoencoder (CAE & QAE).

A. CREDIT CARD FRAUD DETECTION

Legitimate and fraudulent credit card transactions often have similar profiles. Fraudsters constantly adapt to mimic legitimate spending behavior, increasing the difficulty of distinguishing between normal and fraudulent transactions. This results in a highly imbalanced distribution toward legitimate transactions, complicating fraud detection [1]. Machine learning techniques, including supervised, unsupervised, and semi-supervised methods, are the primary approaches to fraud identification [14], [15]. Scalable machine learning algorithms include association rules, fuzzy systems, decision trees, genetic algorithms, neural networks (NNs), SVMs, artificial immune systems, K-nearest neighbor algorithms, and AEs [16] are being used in this manner.

B. CLASSICAL AUTOENCODER (CAE)

The classical autoencoder (CAE) is a pivotal architecture for nearly lossless compression in machine learning. As shown in Figure 1, the autoencoder comprises the primary components including the high-dimensional input data X, an encoder-decoder mechanism, and the reconstructed output data $\hat{X}$. In the encoder-decoder mechanism, the encoder reduces the data to a compact latent space z (the bottleneck), representing the data in its most essential form. This process involves a series of neural network layers mirrored by the decoder, which reconstructs the input data from its compressed state. Suppose that given dataset $\chi = x_i \mid x_i \in \mathbb{R}^n, i = 1, \ldots, N$, the encoder: $\mathbb{R}^n \to \mathbb{R}^k$ with $k \leq n$, compresses each data point into k-dimensional space z. The decoder: $\mathbb{R}^k \to \mathbb{R}^n$ maps z back to the reconstructed $\hat{X}$. The effectiveness of this architecture is measured by how closely the reconstructed data resembles the original. The objective is to minimize loss, which is typically expressed by mean squared error loss functions or binary cross-entropy. These loss functions can be optimized using methods such as gradient descent or stochastic algorithms [17], [18], [19]. This framework is fundamental in applications such as information retrieval [20], feature extraction [21], and anomaly detection [3].

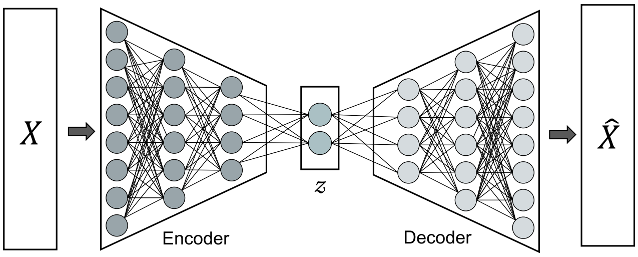

FIGURE 1. CAE architecture. Its primary components are as follows: the high-dimensional input data X; an encoder-decoder mechanism, which compresses the input data into a latent space z, and then reconstructs the data from this compressed form; and the reconstructed output data $\hat{X}$. This process minimizes reconstruction error, making CAEs useful for tasks such as feature extraction, image denoising, and anomaly detection, where such error is represented by the mean squared error loss function or binary cross-entropy.

C. QUANTUM AUTOENCODER (QAE)

QAEs are the quantum analog of CAEs, but the data and operations are quantum mechanical [22]. In a QAE, the quantum encoder (or decoder) is established from a variational quantum circuit [23], [24]. Similar to a CAE, a QAE reduces dimensionality by compressing quantum data into a smaller number of qubits than the input qubits. Thus, the trainable unitary of the quantum encoder part $\mathcal{E}(\theta)$ is defined as $\mathcal{E}(\theta) = \prod_{l=1}^{L} \mathcal{E}_l(\theta)$, where $\mathcal{E}_l(\theta)$ represents a circuit layer composed of a sequence of parameterized single- and two-qubit gates. The parameter θ is tunable to compress the data into the latent space. In a QAE, the set of qubits where data are compressed is called latent qubits, and the qubits that are traced out after this compression are called trash qubits. As depicted in Figure 2, a QAE circuit consists of two trash qubits and the two remaining qubits for latent space preparation. The decoder $\mathcal{D}(\theta)$ applies $\mathcal{E}(\theta)^\dagger$ to the latent and reference qubits to reconstruct the quantum state. Remarkably, the variational quantum circuit in a QAE is trained by minimizing a cost function based on the fidelity or Hamming distance between the trash qubits and the reference state where all reference qubits are in the $|0\rangle$ state [12].

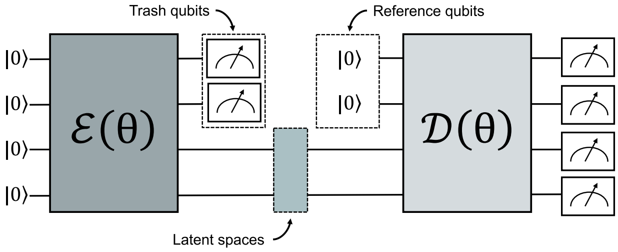

FIGURE 2. QAE architecture. A QAE prepares an initial state of qubits in the $|0\rangle$ state. This is processed through an encoder $\mathcal{E}(\theta)$ and a decoder $\mathcal{D}(\theta)$. The encoder compresses the qubit state into a lower-dimensional quantum space (latent space), and the unused (trash) qubits are traced out. The decoder attempts to reconstruct the original quantum state using both the compressed state and additional (reference) qubits, highlighting the QAE's capability to reduce quantum data dimensions while preserving essential information.

III. RELATED WORK

Fraud detection involves monitoring the behavior of users to estimate, detect, or avoid undesirable behavior [25], [26]. To counter credit card fraud effectively, we begin by shifting our inquiry from the domain of conventional fraud detection to quantum techniques. In classical computing, CAE methods are used to detect credit card fraud. A previous study presented the handling of a highly unbalanced dataset consisting of only 492 fraud transactions and 284,315 genuine transactions [27]. A CAE was compared with logistic regression based on various settings with several thresholds in scenarios involving balanced and unbalanced datasets. The result verified that the CAE was more compatible with the unbalanced dataset. Nevertheless, the advent of quantum computing brings forth quantum machine learning (QML) which provides a computational advantage over the classical approach [28], [29]. QML can significantly transform the paradigm of and approach to classical machine learning by enabling the discovery of novel algorithms that are more efficient than their classical equivalents, especially in fraud detection [30]. For instance, [31] proposed a fraud detection model based on a kernel-based approach with an unsupervised learning model (OC-SVM). An OC-SVM was trained using a subsample dataset to make it manageable for near-term quantum simulations. The model required 20 qubits to reach an average precision of 15%. However, in the current noisy intermediate-scale quantum (NISQ) era, overall systems are more prone to error at larger numbers of qubits [32]. Another quantum credit card fraud detection model was proposed using quantum graph NNs (QGNNs) [33]. In this work, the original transaction features were transformed into graph features through extraction in each batch and epoch. Afterward, a variational quantum classifier was used to help classify graph features. The QGNN model demonstrated a roughly 3% improvement over classical graph NNs across various metrics. However, its recall, a model-sensitive metric used to demonstrate the ability of a model to correctly identify all actual instances as fraudulent transactions, was lower by 10% compared with that of the classical model proposed in [27]. By addressing the challenges of current quantum NN models and capitalizing on the opportunities in the NISQ era, we introduce a novel model that outperforms existing ones. As credit card fraud detection is an anomaly detection problem, we propose a more effective, resource-efficient model. The developed model is based on a QAE, where a parameterized quantum circuit uses only the encoder $\mathcal{E}(\theta)$ component, as depicted in Figure 2, for credit fraud detection. The construction of the model and an in-depth analysis of its performance are thoroughly detailed in the following section.

IV. PROPOSED METHOD

In this section, we provide insights into our proposed hybrid quantum-classical model as QAE-FD. We discuss the overall architecture of QAE-FD, followed by its three main procedures, from data preprocessing to threshold assessment.

A. OVERALL ARCHITECTURE

As depicted in Figure 3, we examine the potential of the QAE algorithm in detecting fraudulent cases from a vast amount of transaction records. We target a highly imbalanced dataset to mirror real-world scenarios [34]. In our approach, the dataset is initially preprocessed to optimize the QML model performance. After the relevant data are preprocessed and normalized, we leverage all classical data in the QAE-FD process for training. In the QAE-FD circuit, the architecture consists of two primary blocks: angle encoding, where we adapt classical datasets into a quantum state, followed by the L layers of the trainable circuit. Each trainable layer comprises a sequence of R_x, R_y, and R_z gates applied on all qubits, followed by a series of Controlled-NOT (CNOT) gates connecting neighboring qubits. Upon completion of the final layer, a designated trash qubit is measured on a computational basis. A loss function of $C(\theta)$ is computed from the measurement outcome of the trash qubits, reflecting how well the quantum circuit has compressed the input data. Alternatively, because one trash qubit is always $|0\rangle$ with a high probability, the minimization of the loss function $C(\theta)$ can be described as the compression of the input from four to three qubits. The trained model is expected to compress genuine data, not fraud data, efficiently. A threshold assessment is performed to determine whether the transaction data whether of high or low loss. The transactions that result in higher fidelity values are classified as fraudulent, whereas those with lower fidelity are flagged as potentially genuine transaction data.

B. DATA PREPROCESSING

As we aim to reflect real-world scenarios, handling a highly imbalanced dataset is a crucial step in this work. This process involves finding, filling, and/or removing null values, standardizing the columns for analysis convenience, and removing duplicate entries in the dataset; this is achieved through principal component analysis (PCA) [35]. Next, we sample the credit card dataset by using a random under-sampling method which is discussed [36], where D is a dataset with K instances. Given its simplicity $D = \{(x_i, y_i)\}_{i=1}^{K}$ where x_i represents the features of the i^{th} instance and y_i represents the class label. Thus, two main categories are considered where the majority of transactions are genuine where N_{genuine} and N_{fraud} are the numbers of instances of genuine and fraudulent transaction data, respectively. The goal is to reduce N_{genuine} to N'_{genuine}, where typically $N'_{\text{genuine}} \approx N_{\text{fraud}}$, to balance the class distribution. Additionally, N'_{genuine} are randomly selected from the numerous amounts of genuine transactions without replacement.

C. QUANTUM AUTOENCODER-BASED FRAUD DETECTION

In this section, we describe our hybrid quantum-classical ansatz circuit, which offers significant advantages for the QAE. As the loss function is determined by the expected measurement values of the qubits, a training circuit that uses only the encoder part is necessary. In this phase, the process begins with the design of an invertible unitary circuit acting upon the initial state $|0\rangle$.

1) ANGLE ENCODING WITHIN QUANTUM CIRCUIT

The normalized transaction data are encoded into a quantum state through angle encoding. In the angle encoding scheme, a feature vector of length n requires $\mathcal{O}(n)$ gates with an n number of qubits [37]. As shown in Figure 3, we set up the state by encoding real-value observable $|\varphi\rangle_i$ as rotation angles along the x axis of the Bloch sphere which can be mathematically represented as

$$|\varphi\rangle = \bigotimes_{i=1}^{n} R_x\left(x_i\right)|\varphi_0\rangle = \bigotimes_{i=1}^{n}\left(\cos\frac{\varphi_i}{2}|0\rangle - i\sin\frac{\varphi_i}{2}|1\rangle\right), \tag{1}$$

where $R_x = e^{-i\psi_i\hat{\sigma}/2}$ is the rotation matrix of each qubit.

2) INTEGRATION OF UNITARY GATE AND LOSS CALCULATION

After the transaction data are encoded into a quantum state, we design the ansatz circuit by applying multivariational ansatz layers. Figure 3 depicts the ansatz, which is utilized as the calibration of transaction dataset exploits on gates R_x, R_y, R_z, and *CNOT* operation while adopting R_x rotation

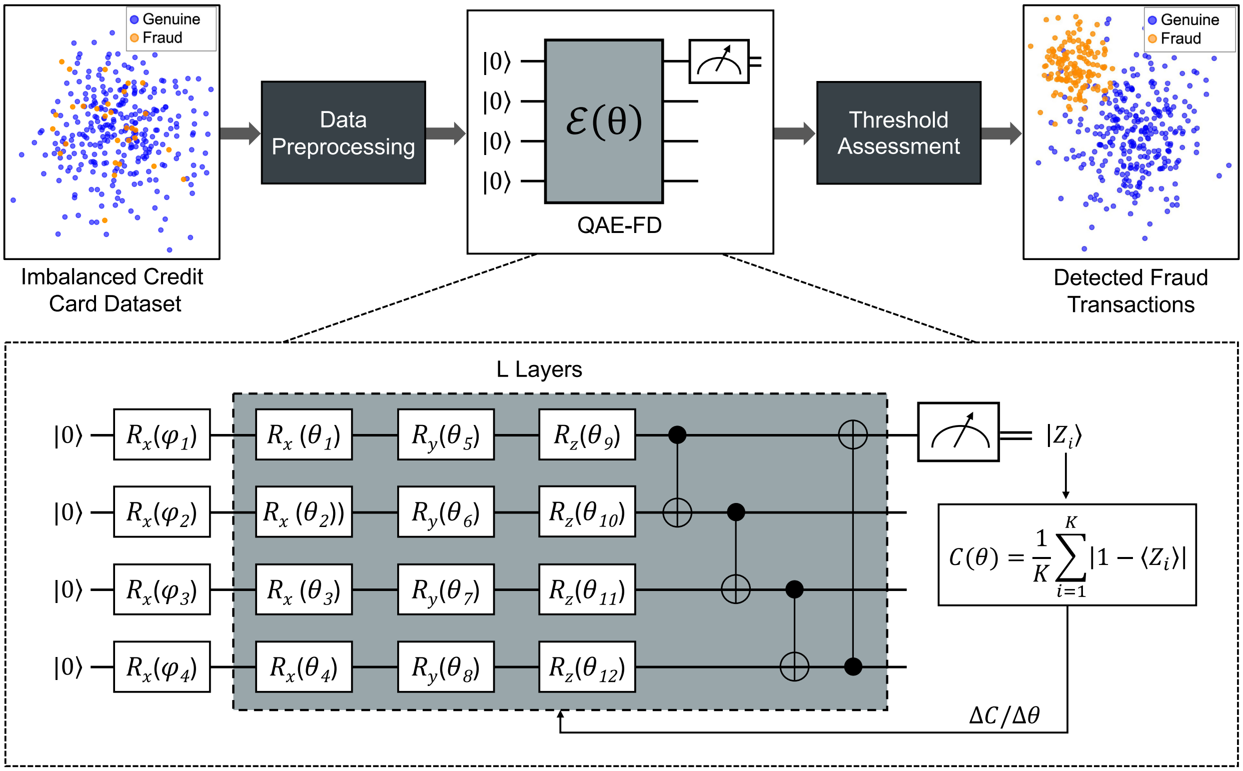

FIGURE 3. Overall workflow of QAE-FD. The workflow commences with the extraction of features from the highly unbalanced credit card dataset, which is then encoded into a quantum Hilbert space to prepare for quantum model training. Subsequently, QAE-FD uses the trash qubits to compute the loss. During training, the parameters of the variational layers are adjusted such that the trash qubits are disentangled from the remaining qubits and kept at $|0\rangle$ while the training information on the unmeasured qubits is preserved simultaneously. Afterward, fraudulent transactions are detected by evaluating the fidelity of the quantum state against a preset threshold in the threshold assessment step. Transactions whose probabilities are greater than or equal to this threshold are classified as genuine, and those with probabilities below the threshold are considered fraudulent.

for angle encoding, as described in the Section IV-C1. Following the final layer, a specified number of n_t trash qubits are measured on a computational basis. This procedure is designed to decouple the trash qubits from the rest of the system effectively by compressing the original ground state into a smaller qubit number. Similarly, we aim to bypass this method as a QAE-based composite quantum system AB as

$$\mathcal{E}(\theta)\,|\psi_{AB}\rangle = |\psi_A\rangle \otimes |\mathrm{trash}\rangle_B, \tag{2}$$

where $|\psi_{AB}\rangle$ denotes the composite system AB. As we concentrate solely on the encoder part, the encoder function compresses the state $|\psi\rangle_{AB}$ of the composite system AB into $|\psi_A\rangle$, exclusively involving subsystem A. Concurrently, it maps subsystem B to a predetermined reference state, referred to as the trash state. The initial quantum state $|\psi_{AB}\rangle$ is obtained using the angle encoding, as described in the previous step. Furthermore, for a given input data, the variational parameters θ of the encoder part are then optimized in order to rotate the trash qubits as closely as possible to the target trash state, where we set $|trash\rangle_B = |0\rangle^{\otimes|B|}$. For a set of the classical transaction data which is used to be analyzed $\mathcal{D} = \{x_i | x_i \in \mathbb{R}^N, i = 1, \ldots K\}$, the QAE is trained using the quantum state obtains as $\mathcal{C} = \{|\psi_x\rangle \mid \forall x \in \mathcal{D}\}$, where $|\psi_x\rangle$ is the encoding state using angle encoding. Hence, the loss function $\mathcal{C}(\theta)$ that measures the reconstruction error is achieved by mean of the training procedure whose aims to find the optimal parameter θ^* as

$$\mathcal{C}(\theta) = \frac{1}{K}\sum_{i=1}^{K}|1 - \langle Z_i\rangle| \tag{3}$$

where Z_i is the expectation value of the Pauli-Z operator on the trash qubit corresponding to the i^{th} transaction. The loss function is faithful, which means that it reaches the global minimum $\mathcal{C}(\theta^*) = 0$ only when $\langle Z_i\rangle = 1, \forall i = 1, \ldots, K$, which is when the trash qubit is always and perfectly disentangled from the other qubits and mapped to the target trash state $|0\rangle$.

D. THRESHOLD ASSESSMENT

After the QAE is trained to learn the compressed representation of the original information, the compressed quantum state is used as input for classification. Label assignment is conducted based on a majority vote on multiple shots of the same quantum circuit: input of the transaction dataset is set to "genuine" if most measurements give $|0\rangle$ as an outcome, and "fraudulent" otherwise. Formally, let be

$\rho_x^G = \mathrm{Tr}_F\left[\mathcal{E}(\theta)(|\psi_x\rangle\langle\psi_x|)\mathcal{E}(\theta)^\dagger\right]$ and $\mathcal{T}$ be a predefined threshold. Then, the label is assigned according to the following decision rule:

$$\hat{y}_i = \begin{cases} 0, & \text{if } p_0 = \mathrm{Tr}(|0\rangle\langle 0|\, U^G_{\rho_{x_i}} U^\dagger) \geq \mathcal{T}, \\ 1, & \text{otherwise} \end{cases} \tag{4}$$

where ρ_0 denotes the probability that the measurement yields $|0\rangle$ as an outcome and U represents for the variational unitary results from the $U(\alpha, \gamma)$. More than that, the loss function used to classify the transaction is the cross-entropy, which can be optimized using the classical optimizer in [38].

V. EVALUATION

In this section, we assess our proposed model's performance. We start with the preparation of our experimental setup and then evaluate the performance of our proposed method.

A. EXPERIMENTAL SETUP

We present the experimental setup for evaluation, specifically the datasets, hyperparameter, model configuration, and evaluation metrics used to assess our proposed model.

1) DATASET PREPARATION

We use the publicly available credit card transaction dataset of European cardholders from [34] in our experiment. This dataset includes 30 features along with the output feature, which labels transactions as genuine (0) or fraudulent (1). The original feature labels except "Time" and "Amount" are omitted. The retained labels are named from $V1$ to $V28$ due to privacy concerns, corresponding to the output of the PCA transformation output. The "Time" feature records the number of seconds between each transaction and the dataset's first transaction. The "Amount" feature represents the amount of money involved, we visualized the presentation in Figure 4, which represents the correlation matrix used to show the strength of interaction and influence between the features. A positive correlation indicates that an increase in one feature's value leads to an increase in another feature's value, while a negative correlation indicates the opposite. No correlation suggests the features are independent of each other. This process is implemented alongside random under-sampling using the Python library called Scikit-learn and Imbalanced-learn [39], [40].

2) HYPERPARAMETER AND MODEL CONFIGURATION

The quantum circuit configuration for the QAE-FD model consists of a variational quantum circuit designed to compress input data into a lower-dimensional quantum state. The circuit employs four qubits, with a variational depth of four layers, each comprising parameterized single-qubit rotation gates R_x, R_y, and R_z followed by two-qubit Controlled-NOT (CNOT) gates for entanglement. Threshold values, set empirically at 0.5, 0.6, and 0.7, are used for anomaly detection, where each threshold reflects varying levels of sensitivity in distinguishing between genuine and fraudulent transactions. These thresholds were chosen through validation experiments, balancing false positives and false negatives while maximizing detection accuracy. The circuit configuration ensures efficient data compression and anomaly detection by leveraging quantum principles.

Table 1 shows the hyperparameter configuration used to optimize the QAE for accurate, efficient fraud detection. Setting the number of epochs to 50, the batch size to 16, and the learning rate to 0.001 ensures thorough, balanced training. The Adam optimizer helps in dealing with the complexities of transaction data while using 4 ansatz layers with 4 qubits to exploit quantum advantages for better pattern recognition. There are three threshold parameters that fine-tune the model's sensitivity to fraud detection, and the Pennylane framework [41] facilitates the seamless integration of quantum and classical machine learning techniques.

TABLE 1. Hyperparameter and model configuration. The table lists the critical values for various hyperparameters which include the number of epochs, batch size, learning rate, optimizer, number of ansatz layers, thresholds, number of qubits, and the framework that is employed for architecture implementation.

Hyperparameter	Value
Number of epochs	50
Batch size	16
Learning rate	0.001
Optimizer	Adam
Number of ansatz layers	4
Threshold	0.5, 0.6, & 0.7
Number of qubits	4
Framework for architecture implementation	Pennylane

3) EVALUATION METRIC

To demonstrate the effectiveness of our proposed method, we present the evaluation metric needed to evaluate our method in quantum machine learning.

a: CONFUSION MATRIX

It is used to describe the performance of the proposed model for selecting a dataset and is in the form of four different sets of expected real values, where the confusion matrix provides the number of transactions per set.

- True Positive (TP): denotes as the number of fraudulent transactions that the model correctly identified as fraudulent transactions.
- False Positive (FP): denotes as the number of genuine transactions that the model incorrectly identified as fraudulent transactions.
- False Negative (FN): denotes as the number of fraudulent transactions that the model incorrectly identified as genuine transactions.
- True Negative (TN): denotes as the number of genuine transactions that the model correctly identified as genuine transactions.

b: OTHER CRITICAL METRICS

These are standard metrics for unbalanced datasets.

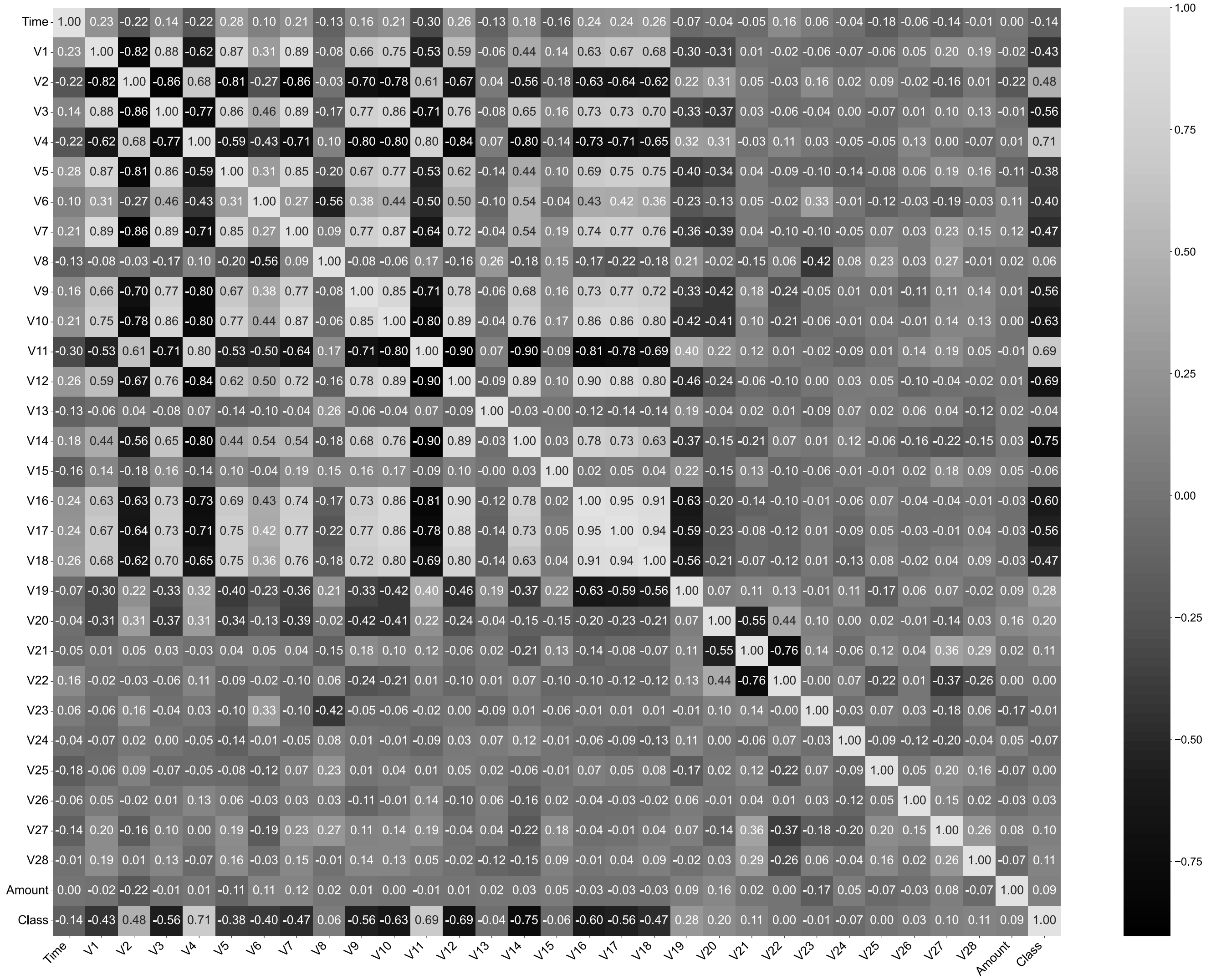

FIGURE 4. Correlation matrix of the credit card dataset [34]. This dataset encompasses encrypted features *V*1 through *V*28 with Time and Amount. The correlation matrix is utilized to analyze the relationship between each feature. A positive correlation indicates that an increase in one feature's value corresponds to an increase in another feature's value, whereas a negative correlation suggests the opposite. When there is no correlation, it means the features are independent of each other.

Precision is the ratio of fraudulent transactions (TP) to the total predicted fraudulent transactions (TP and FP). It indicates the accuracy of fraud predictions among those labeled as fraudulent.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{5}$$

Recall (sensitivity) is the ratio of fraudulent transactions (TP) to all actual fraudulent transactions (TP and FN). It measures the model's ability to identify all instances of fraud, capturing the effectiveness of the model in detecting frauds that actually occurred.

$$\text{Recall} = \frac{TP}{TP + FN} \tag{6}$$

Accuracy is the ratio of correctly predicted fraudulent and genuine transactions to the total transactions. It measures the overall correctness of the model.

$$\text{Accuracy} = \frac{TN + TP}{TN + TP + FN + FP} \tag{7}$$

F1-score is the harmonic mean of Precision and Recall. It provides a single metric that balances both the concerns of false positives and false negatives. It is particularly useful when the class distribution is imbalanced.

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{8}$$

Specificity evaluates how well a classification model is at identifying negative examples. It is particularly important in scenarios where the cost of FP is high.

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{9}$$

G-mean is a metric that balances recall and specificity. It is particularly useful when you want to assess a model's ability to perform well across both classes (positive and negative) in datasets where there might be a class imbalance.

$$\text{G-mean} = \sqrt{\text{Sensitivity} \times \text{Specificity}} \tag{10}$$

AUC is the area under the receiver operating characteristic curve. It is a single scalar value that measures the overall performance of a model across all threshold values. The AUC helps quantify how well a model distinguishes between classes (e.g., fraudulent versus genuine).

B. MODEL PERFORMANCE

In quantum computing, we illustrate both noise-free and noisy environments with various performance matrices that are crucial for evaluating the effectiveness of our algorithm and model.

1) NOISE-FREE ENVIRONMENT

To present the model performance, we first illustrate a noise-free environment on the simulator with no noise from the confusion matrix and other critical matrices.

- **Confusion matrix:** Confusion matrices are essential for evaluating the performance of our proposed model, particularly in classification tasks, such as distinguishing between genuine and fraudulent transactions. Table 2 provides a detailed breakdown of the performance metrics of our model by displaying the counts of TN, FP, FN, and TP at three different threshold levels ($\mathcal{T}$ = 0.5, $\mathcal{T}$ = 0.6, and $\mathcal{T}$ = 0.7). At $\mathcal{T}$ = 0.5, the model correctly identified 55,953 transactions as genuine (TN) but misclassified 911 as fraudulent (FP). It also successfully detected 91 instances of fraudulent cases (TP) but missed 7 fraudulent transactions (FN). When the threshold was increased to $\mathcal{T}$ = 0.6, there was a notable improvement in identifying genuine transactions where the TN count increased to 56,430, and the FP count decreased significantly to 434. However, this adjustment also led to a slight increase in missed fraud cases, with the FN count rising to 13, and the TP count decreased in correctly identified fraud cases to 85. At $\mathcal{T}$ = 0.7, the TN increases to 56,701 due to the reduction in FP count to 163. Interestingly, this higher threshold setting also resulted in a marginal decrease in missed fraud cases, with FN dropping to 10, and a slight increase in accurate fraud detection, with TP rising to 88. This progression illustrates a trade-off between detecting fraud more accurately and increasing the chance of missing fraudulent transactions (higher FN). Each threshold adjustment shows a balance between minimizing false alarms and capturing true fraudulent cases. This accurate measurement is crucial in quantum computing, where the inherent advantages of quantum algorithms such as faster processing compared to classical algorithms can be leveraged to significantly enhance the detection mechanisms in scenarios such as real-time financial fraud detection.

TABLE 2. Confusion matrix on a noise-free environment on the different threshold number. The table demonstrates the obtained results including TN, FN, FP, and TP respectively from various threshold $\mathcal{T}$ ranging from 0.5 to 0.7.

Threshold	Transaction	Genuine	Fraud
$\mathcal{T}$ = 0.5	Genuine	TN = 55,953	FP = 911
	Fraud	FN = 7	TP = 91
$\mathcal{T}$ = 0.6	Genuine	TN = 56,430	FP = 434
	Fraud	FN = 13	TP = 85
$\mathcal{T}$ = 0.7	Genuine	TN = 56,701	FP =163
	Fraud	FN = 10	TP = 88

- **Other critical metrics:** Table 3 provides a detailed assessment of various performance metrics for the fraud detection model in the noise-free quantum computing environment across the three thresholds ($\mathcal{T}$ = 0.5, $\mathcal{T}$ = 0.6, and $\mathcal{T}$ = 0.7). These metrics are precision, recall, accuracy, F1-score, specificity, G-mean, and AUC. At the lowest threshold of $\mathcal{T}$ = 0.5, the model achieves high recall (0.928), indicating effective identification of fraudulent transactions, but has low precision (0.090), resulting in a significant number of FPs. The F1-score at this threshold is 0.170, reflecting the imbalance between precision and recall. As the threshold increases to $\mathcal{T}$ = 0.6, both precision and recall improve slightly, leading to a better F1-score of 0.275. The highest threshold, $\mathcal{T}$ = 0.7, shows a marked improvement in precision (0.379) and a robust recall (0.897), yielding the highest F1-score of 0.533. This indicates a balanced trade-off between detecting fraud and minimizing false alarms. Specificity is high across all thresholds, showing that the model proficiently identifies genuine transactions. Along with accuracy is performing better in every threshold which indicates the correctness of the overall model. The G-mean and AUC metrics illustrate the model's overall effectiveness and its capability to distinguish between fraudulent and genuine transactions at different operational thresholds.

2) NOISY ENVIRONMENT

To study the actual performance of the model, we experiment using a noisy environment of the 27-qubit fake backend FakeCairo (IBM) with the same confusion matrix and other critical metrics.

- **Confusion matrix:** The confusion matrix in Table 4 offers insights into the model's performance in distinguishing between genuine and fraudulent cases at the threshold setting of 0.7 in the noisy environment. The matrix showcases the true negatives (TN = 56,783) and false positives (FP = 81) for genuine transactions, alongside the true positives (TP = 75) and false negatives (FN = 23) for fraudulent transactions. These results highlight the model's high accuracy in correctly identifying genuine cases, as evidenced by the substantial number of TNs. However, the presence of FPs (instances where genuine cases are incorrectly

TABLE 3. Other critical metrics on the noise-free environment at different threshold levels. The table depicts the various performance metrics which include precision, recall, accuracy, F1-score, specificity, G-mean, and AUC respectively from various threshold $\mathcal{T}$ ranging from 0.5 to 0.7.

Threshold	Precision	Recall	Accuracy	F1-Score	Specificity	G-mean	AUC
$\mathcal{T} = 0.5$	0.090	0.928	0.980	0.170	0.980	0.955	0.950
$\mathcal{T} = 0.6$	0.163	0.867	0.992	0.275	0.992	0.927	0.929
$\mathcal{T} = 0.7$	0.379	0.897	0.997	0.533	0.997	0.946	0.947

flagged as fraudulent) is a common challenge in noisy environments where data imperfections can lead to misclassifications. Nonetheless, the relatively low number of FNs demonstrates the model's effectiveness in detecting actual fraudulent activities, though the number of TPs suggests it still has room for improvement in sensitivity. Overall, the confusion matrix underscores the model's robustness in handling noise, with a strong emphasis on minimizing FNs to ensure that true fraud cases are not overlooked, and reducing FPs to maintain precision.

TABLE 4. Confusion matrix on noisy environment. The table presents the obtained result from model training under the error-prone noisy environment by detailing TN, FN, FP, and TP at the most sensitive threshold level $\mathcal{T} = 0.7$.

Transaction	Genuine	Fraud
Genuine	TN = 56,783	FP = 81
Fraud	FN = 23	TP = 75

- **Other critical metrics:** The performance of the proposed QAE-FD should be evaluated using other critical metrics in a noisy environment to understand its robustness and efficacy. Table 5 demonstrates the performance of our proposed model at the threshold value of 0.7. The precision of 0.480, while lower than the recall, reflects the model's ability to correctly identify fraud cases out of all cases flagged as fraud. The recall rate of 0.765 suggests that our QAE is effective at identifying a significant portion of actual fraud cases, which is crucial in minimizing undetected fraudulent activities. The model achieves an impressive accuracy of 99%, indicating its high capability to classify both fraudulent and genuine cases correctly. Its F1-score of 0.590 suggests a balanced measure of the model's precision and recall, emphasizing its overall performance in handling noisy data. Its specificity is 0.881, confirming its strength in correctly identifying genuine transactions. The G-mean, which is a geometric mean of recall and specificity and provides an overall measure of the model's performance across both classes, is 0.874. Lastly, the AUC is 0.881, illustrating the strong discriminative ability of our model to distinguish between different classes. The high values of recall, F1-score, and other metrics highlight the model's effectiveness in ensuring that fraudulent activities are not overlooked, even amidst noise and data imperfections.

C. MODEL COMPARISON

Table 6 highlights the superior performance of our proposed QAE-FD compared with a CAE [27], Quantum One-Class Support Vector Machine (QO-SVM) [31], and Quantum Graph Neural Network (QGNN) [33], highlighting its enhanced capabilities across various metrics.

In the context of fraud detection, the QAE-FD model significantly outperforms the CAE. Although the CAE has a high recall of 0.917, its precision is only 0.090, indicating a large number of FPs; thus, it is less reliable in practical applications. By contrast, QAE-FD maintains a much better balance, with a precision of 0.379 and a superior recall of 0.897; therefore, it not only captures most fraudulent transactions but also maintains a lower rate of false alarms. Additionally, QAE-FD's accuracy of 0.990 far surpasses CAE's 0.800, and its AUC (a metric unavailable for the CAE) of 0.947 indicates its robust ability to discriminate between fraudulent and genuine transactions.

Comparing QAE-FD with QO-SVM in fraud detection reveals the strengths of the proposed model. Although QO-SVM achieves a commendable precision of 0.700, it lacks available data for recall and accuracy, limiting our assessment of its effectiveness. On the contrary, QAE-FD not only provides all necessary metrics but also excels in recall (0.897) and accuracy (0.990), highlighting its reliability and robustness in fraud detection. QAE-FD's AUC of 0.947 further underscores its superior classification capabilities compared with QO-SVM (unreported AUC). Additionally, QAE-FD's efficiency is evident; it requires only 4 qubits, showcasing better quantum resource management than QO-SVM, which needs 20 qubits.

Compared with the QGNN, QAE-FD again demonstrates superiority in fraud detection. Although the QGNN shows an impressive precision of 0.945, its recall (0.795) is lower than that of QAE-FD. Thus, QAE-FD is more effective in identifying TPs. Furthermore, QAE-FD's accuracy of 0.990 surpasses QGNN's 0.920, and its AUC of 0.947 is considerably higher than QGNN's unreported AUC, affirming its enhanced ability to differentiate between fraudulent and genuine transactions accurately. Requiring only 4 qubits compared with the QGNN's 6, QAE-FD also highlights its efficiency in using quantum resources.

Overall, the proposed QAE-FD model stands out as the most effective and efficient model for fraud detection compared with the CAE, QO-SVM, and QGNN. It not only achieves higher accuracy and recall but also uses quantum resources more efficiently than the other models. Hence, it is a highly effective and efficient solution to fraud detection in noisy quantum computing environments.

VI. DISCUSSION AND LIMITATION

Our proposed QAE-FD performs robustly across various metrics; however, it does have limitations, particularly in

TABLE 5. Other critical metrics on the noisy environment. This table exhibits the model's performance under simulated noisy conditions, including precision, recall, accuracy, F1-score, specificity, G-mean, and AUC. The results are shown for the most sensitive scenario with a threshold of $\mathcal{T} = 0.7$.

Precision	Recall	Accuracy	F1-Score	Specificity	G-mean	AUC
0.480	0.765	0.998	0.590	0.881	0.874	0.881

TABLE 6. Model comparison between classical, existing quantum, and our proposed methods. This table provides a detailed comparison between our proposed model and existing classical and quantum models. Our model, QAE-FD, with an overall AUC performance score of 0.947, achieves superior precision and accuracy compared to classical models by effectively reducing the false alarm rate in fraud detection. Additionally, it demonstrates enhanced performance in a critical metric, recall. This improvement signifies that our model can accurately classify the actual number of fraud cases in a highly imbalanced dataset while efficiently utilizing only 4 qubits.

Model	Precision	Recall	Accuracy	F1-Score	AUC	Qubit
CAE	0.090	0.917	0.800	N/A	N/A	N/A
QO-SVM	0.700	N/A	N/A	N/A	N/A	20
QGNN	0.945	0.795	0.920	0.860	N/A	6
QAE-FD	0.379	0.897	0.990	0.533	0.947	4

TABLE 7. Ablation study on the effect of circuit layer (L) in QAE-FD model. Results indicate an optimal balance between fraud detection effectiveness and false alarm minimization is achieved with $L = 3$ to $L = 5$ layers. Increasing layers beyond this range do not correspond to linear improvements, potentially leading to diminishing returns and performance degradation due to overfitting and difficulties in generalizing across imbalanced datasets.

Layer	Precision	Recall	Accuracy	F1-Score	Specificity	G-mean	AUC
L = 1	0.000	0.000	0.990	0.024	1.000	0.000	0.500
L = 2	0.500	0.816	0.990	0.583	0.998	0.902	0.907
L = 3	0.379	0.897	0.990	0.533	0.997	0.946	0.947
L = 4	0.350	0.897	0.990	0.504	0.997	0.946	0.947
L = 5	0.426	0.800	0.990	0.555	0.998	0.891	0.897
L = 6	0.362	0.836	0.990	0.506	0.997	0.913	0.917

terms of sensitivity. Although our model achieves commendable accuracy and specificity, its sensitivity, as reflected in its precision and recall values, reveals certain constraints when handling noisy data and complex fraud scenarios. These limitations necessitate a deeper exploration of how adjustments in the quantum circuit's complexity and number of layers will affect the overall model performance.

Table 7 presents a nuanced exploration of how changes in the number of layers (L) in the quantum circuit affect the performance of QAE-FD. Specifically, the table shows how different model configurations (and thus levels of complexity) affect precision, recall, accuracy, F1-score, specificity, G-mean, and AUC, leading to a discussion on the trade-offs and limitations associated with each configuration.

The model with the simplest configuration $L = 1$ cannot effectively detect fraudulent transactions, as evidenced by precision and recall values of 0 despite the 0.990 accuracy. This indicates that the model is overly conservative in this configuration, primarily classifying most transactions as genuine, hence the specificity of 1.000. This configuration results in a high number of FNs; fraudulent activities are not detected, as reflected by the very low F1-score and AUC value of 0.500.

$L = 2$ significantly enhances the model's ability to detect fraud, with its recall improving drastically to 0.816, although precision remains moderate at 0.500. This layer configuration begins to address the sensitivity limitations observed in the single-layer model, balancing the detection of fraudulent transactions and a manageable rate of FPs. Its accuracy remains high (0.990), its specificity slightly declines but remains robust, and its AUC improves to 0.907, indicating better model balance.

The $L = 3$ configuration optimizes the trade-off between precision and recall, with values of 0.379 and 0.897, respectively. This layer setup not only maintains high accuracy (0.990) but also achieves a good F1-score (0.533). The high G-mean and AUC of 0.946 and 0.947 further show the model's effective discrimination between fraudulent and genuine transactions. This configuration is the most balanced in terms of overall model performance.

The $L = 4$ and $L = 5$ configurations show continuous adjustments in the balance between model sensitivity and specificity. Both configurations exhibit slight variations in precision and recall, with $L = 4$ slightly outperforming $L = 5$ in G-mean and AUC. These configurations indicate that further increases in model complexity may begin to yield diminishing returns (i.e., declining model efficiency in credit card fraud detection).

Layer $L = 6$ demonstrates a potential onset of overfitting or unnecessary complexity, as evidenced by a slight regression in the performance metrics, such as precision, F1-score, and AUC. Thus, adding layers beyond this point may not necessarily enhance the model's capability and can compromise its ability to generalize effectively across different scenarios.

This detailed evaluation reveals that although QAE-FD is effective, the design of the quantum circuit for fraud detection has a critical limitation: increasing the number of layers does not linearly improve model performance across all metrics. Instead, it suggests an optimal midpoint (approximately

$L = 3$ to $L = 5$), where the model achieves the best balance between detecting fraud and minimizing false alarms. Beyond this point, additional layers may lead to diminishing returns or even degrade performance, illustrating the complex interplay between circuit depth, model overfitting, and generalization ability across noisy data. This analysis is crucial for guiding architectural decisions in QML models, particularly in balancing complexity with the real-world practical efficacy.

VII. CONCLUSION

This study substantiates the considerable potential of QAEs in augmenting the detection of fraudulent activities within imbalanced credit card transaction datasets. The innovative application of quantum computing principles in our QAE-FD model not only markedly enhances the detection rates of fraudulent transactions but also significantly expedites analytical processing. This dual improvement addresses pivotal challenges inherent in the anomaly detection systems traditionally used in financial sectors. The core of our approach is the use of quantum mechanical properties to compress high-dimensional transaction data into lower-dimensional quantum states, enhancing the system's ability to detect subtle anomalies indicative of fraud. The methodology has a bifurcated operational structure: transforming classical transaction data into quantum states and then applying quantum classification techniques. This novel strategy establishes new benchmarks for accuracy and computational efficiency in fraud detection mechanisms. Empirical validation performed on a real-world dataset corroborates the enhanced performance of the QAE-FD model over conventional machine learning models, thereby highlighting the operational feasibility of quantum approaches in real-time settings. Our findings prompt the expanded exploration and integration of quantum technologies in financial security frameworks, suggesting the significant potential of these technologies to advance frontline defenses against credit card fraud. Furthermore, this study contributes to both theoretical and practical enhancements in the QML field. It paves the way for future scholarly research on the integration of quantum computing technologies across various aspects of anomaly detection and cybersecurity using different simulators and real quantum computers, such as Qiskit (IBM), Cirq (Google), and Forest (Rigetti). Given the scalability of the proposed model, it shows substantial potential for real-world application beyond the financial sector, potentially revolutionizing approaches to data anomaly detection in an increasingly digitized global landscape. Future research will prioritize a systematic evaluation of the QAE-FD model in comparison with classical machine learning algorithms to establish a comprehensive performance benchmark such as Logistic Regression, Random Forest, and Support Vector Machines. A comparative study of computational resource requirements will yield critical insights into the model's operational efficiency like execution time, memory consumption, and hardware dependencies. Furthermore, the simulation of a realistic quantum noise model will be undertaken to rigorously assess the model's robustness. Enhancing noise resilience through advanced error mitigation techniques, coupled with the development of interpretability frameworks to clarify latent space representations and circuit dynamics, will further strengthen the model's practical utility in real-world fraud detection scenarios.

CODE AVAILABILITY

The code that supports the findings of this study is openly available in the Github repository, QAE-FD.

REFERENCES

[1] J. O. Awoyemi, A. O. Adetunmbi, and S. A. Oluwadare, "Credit card fraud detection using machine learning techniques: A comparative analysis," in *Proc. Int. Conf. Comput. Netw. Informat. (ICCNI)*, Oct. 2017, pp. 1–9.

[2] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Comput. Surv.*, vol. 41, no. 3, pp. 1–58, Jul. 2009.

[3] C. Zhou and R. C. Paffenroth, "Anomaly detection with robust deep autoencoders," in *Proc. 23rd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2017, pp. 665–674.

[4] G. Cohen, S. Afshar, J. Tapson, and A. van Schaik, "EMNIST: Extending MNIST to handwritten letters," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, May 2017, pp. 2921–2926.

[5] R. A. Fisher, "Iris," UCI Mach. Learn. Repository, Rep., 1988, doi: 10.24432/C56C76.

[6] A. Dal Pozzolo, O. Caelen, Y.-A. Le Borgne, S. Waterschoot, and G. Bontempi, "Learned lessons in credit card fraud detection from a practitioner perspective," *Expert Syst. Appl.*, vol. 41, no. 10, pp. 4915–4928, Aug. 2014.

[7] N. Liu and P. Rebentrost, "Quantum machine learning for quantum anomaly detection," *Phys. Rev. A, Gen. Phys.*, vol. 97, no. 4, Apr. 2018, Art. no. 042315.

[8] D. Herr, B. Obert, and M. Rosenkranz, "Anomaly detection with variational quantum generative adversarial networks," *Quantum Sci. Technol.*, vol. 6, no. 4, Oct. 2021, Art. no. 045004.

[9] J. Romero, J. P. Olson, and A. Aspuru-Guzik, "Quantum autoencoders for efficient compression of quantum data," *Quantum Sci. Technol.*, vol. 2, no. 4, Dec. 2017, Art. no. 045001.

[10] S. Bordoni, D. Stanev, T. Santantonio, and S. Giagu, "Long-lived particles anomaly detection with parametrized quantum circuits," *Particles*, vol. 6, no. 1, pp. 297–311, Feb. 2023.

[11] K. Kottmann, F. Metz, J. Fraxanet, and N. Baldelli, "Variational quantum anomaly detection: Unsupervised mapping of phase diagrams on a physical quantum computer," *Phys. Rev. Res.*, vol. 3, no. 4, Dec. 2021, Art. no. 043184.

[12] G. Park, J. Huh, and D. K. Park, "Variational quantum one-class classifier," *Mach. Learn., Sci. Technol.*, vol. 4, no. 1, Mar. 2023, Art. no. 015006.

[13] M. Grossi, N. Ibrahim, V. Radescu, R. Loredo, K. Voigt, C. von Altrock, and A. Rudnik, "Mixed quantum–classical method for fraud detection with quantum feature selection," *IEEE Trans. Quantum Eng.*, vol. 3, pp. 1–12, 2022.

[14] M. Rezapour, "Anomaly detection using unsupervised methods: Credit card fraud case study," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 11, 2019.

[15] N. Dzakiyullah, "Semi-supervised classification on credit card fraud detection using AutoEncoders," *J. Appl. Data Sci.*, vol. 2, no. 1, pp. 1–7, Jan. 2021.

[16] S. Misra, S. Thakur, M. Ghosh, and S. K. Saha, "An autoencoder based model for detecting fraudulent credit card transaction," *Proc. Comput. Sci.*, vol. 167, pp. 254–262, Jan. 2020.

[17] M. Sewak, S. K. Sahay, and H. Rathore, "An overview of deep learning architecture of deep neural networks and autoencoders," *J. Comput. Theor. Nanosci.*, vol. 17, no. 1, pp. 182–188, Jan. 2020.

[18] V. S. Ngairangbam, M. Spannowsky, and M. Takeuchi, "Anomaly detection in high-energy physics using a quantum autoencoder," *Phys. Rev. D, Part. Fields*, vol. 105, no. 9, May 2022, Art. no. 095004.

[19] M. Yousefi-Azar, V. Varadharajan, L. Hamey, and U. Tupakula, "Autoencoder-based feature learning for cyber security applications," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, May 2017, pp. 3854–3861.

[20] J. Pfeiffer, S. Broscheit, R. Gemulla, and M. Göschl, "A neural autoencoder approach for document ranking and query refinement in pharmacogenomic information retrieval," in *Proc. BioNLP Workshop*, 2018, pp. 87–97.

[21] Q. Meng, D. Catchpoole, D. Skillicom, and P. J. Kennedy, "Relational autoencoder for feature extraction," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, May 2017, pp. 364–371.

[22] M. Hdaib, S. Rajasegarar, and L. Pan, "Quantum autoencoder frameworks for network anomaly detection," in *Proc. Int. Conf. Neural Inf. Process.* Singapore: Springer, 2023.

[23] M. Cerezo, A. Poremba, L. Cincio, and P. J. Coles, "Variational quantum fidelity estimation," *Quantum*, vol. 4, p. 248, Mar. 2020.

[24] Y. Du and D. Tao, "On exploring the potential of quantum auto-encoder for learning quantum systems," 2021, *arXiv:2106.15432*.

[25] T. P. Bhatla, V. Prabhu, and A. Dua, "Understanding credit card frauds," *Cards Bus. Rev.*, vol. 1, no. 6, pp. 1–15, 2003.

[26] B. Masuda, "Credit card fraud prevention: A successful retail strategy," *Crime Prevention Stud.*, vol. 1, pp. 34–121, 1993.

[27] M. A. Al-Shabi, "Credit card fraud detection using autoencoder model in unbalanced datasets," *J. Adv. Math. Comput. Sci.*, vol. 33, no. 5, pp. 1–16, Aug. 2019.

[28] P. Wittek, *Quantum Machine Learning: What Quantum Computing Means To Data Mining*. New York, NY, USA: Academic, 2014.

[29] S. Lloyd, M. Mohseni, and P. Rebentrost, "Quantum algorithms for supervised and unsupervised machine learning," 2013, *arXiv:1307.0411*.

[30] N. Innan, M. A.-Z. Khan, and M. Bennai, "Financial fraud detection: A comparative study of quantum machine learning models," *Int. J. Quantum Inf.*, vol. 22, no. 2, Mar. 2024, Art. no. 2350044.

[31] O. Kyriienko and E. B. Magnusson, "Unsupervised quantum machine learning for fraud detection," 2022, *arXiv:2208.01203*.

[32] J. Preskill, "Quantum computing in the NISQ era and beyond," *Quantum*, vol. 2, p. 79, Aug. 2018.

[33] N. Innan, A. Sawaika, A. Dhor, S. Dutta, S. Thota, H. Gokal, N. Patel, M. A.-Z. Khan, I. Theodonis, and M. Bennai, "Financial fraud detection using quantum graph neural networks," *Quantum Mach. Intell.*, vol. 6, no. 1, pp. 1–18, Jun. 2024.

[34] ULB Machine Learning Group. (2013). *Credit Card Fraud Detection*. [Online]. Available: https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud

[35] Z. Salekshahrezaee, J. L. Leevy, and T. M. Khoshgoftaar, "Feature extraction for class imbalance using a convolutional autoencoder and data sampling," in *Proc. IEEE 33rd Int. Conf. Tools Artif. Intell. (ICTAI)*, Nov. 2021, pp. 217–223.

[36] T. M. Alam, K. Shaukat, I. A. Hameed, S. Luo, M. U. Sarwar, S. Shabbir, J. Li, and M. Khushi, "An investigation of credit card default prediction in the imbalanced datasets," *IEEE Access*, vol. 8, pp. 201173–201198, 2020.

[37] R. LaRose and B. Coyle, "Robust data encodings for quantum classifiers," *Phys. Rev. A, Gen. Phys.*, vol. 102, no. 3, Sep. 2020, Art. no. 032420.

[38] R. Y. Rubinstein, "The cross-entropy method for combinatorial and continuous optimization," *Methodol. Comput. Appl. Probab.*, vol. 1, no. 2, pp. 127–190, Sep. 1999.

[39] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Nov. 2011.

[40] G. Lemaître, F. Nogueira, and C. K. Aridas, "Imbalanced-learn: A Python toolbox to tackle the curse of imbalanced datasets in machine learning," *J. Mach. Learn. Res.*, vol. 18, no. 17, pp. 1–5, 2017.

[41] V. Bergholm et al., "PennyLane: Automatic differentiation of hybrid quantum-classical computations," 2018, *arXiv:1811.04968*.

CHANSREYNICH HUOT received the bachelor's degree in informatics economics from the Royal University of Law and Economics, Cambodia, in 2020. She is currently pursuing the M.S. degree with the Department of AI Convergence, Pukyong National University, Busan, South Korea. Her research interest includes classical and quantum machine learning.

SOVANMONYNUTH HENG received the bachelor's degree in computer science from Norton University, Phnom Penh, Cambodia, in 2020. She is currently pursuing the M.S. degree with the Department of AI Convergence, Pukyong National University, Busan, South Korea. Her research interests include quantum computing and quantum machine learning.

TAE-KYUNG KIM received the M.E. and Ph.D. degrees from the Department of Information Industry Engineering, Chungbuk National University, Cheongju, Republic of Korea, in 2005 and 2010, respectively. From 2013 to 2019, he was the Director of Korea Software HRD Center, Phnom Penh, Cambodia. He was an Assistant Professor with the AI Department of Computer and Information Technology, Incheon Jaeneung University, Republic of Korea, from 2020 to 2023. Since 2024, he has been an Assistant Professor with the Department of Management Information Systems, Chungbuk National University, Cheongju, South Korea. His research interests include big data, artificial intelligence, and software education.

YOUNGSUN HAN (Member, IEEE) received the B.S. and Ph.D. degrees in electrical engineering from Korea University, Seoul, South Korea, in 2003 and 2009, respectively. He was a Senior Engineer with the System LSI, Samsung Electronics, Suwon, South Korea, from 2009 to 2011. He was an Assistant/Associate Professor with the Department of Electronic Engineering, Kyungil University, Gyeongsan, South Korea, from 2011 to 2019. He is currently a Professor with the Department of Computer Engineering, Pukyong National University, Busan, South Korea. His research interests include quantum computing, compiler construction, microarchitecture, and high-performance computing.