2024 Global Conference on Communications and Information Technologies (GCCIT)
Bangalore, India. Oct 25-26, 2024

2024 Global Conference on Communications and Information Technologies (GCCIT) | 979-8-3503-8891-6/24/$31.00 ©2024 IEEE | DOI: 10.1109/GCCIT63234.2024.10862508

Fraud Detection using Quantum Machine Learning(QML)

Sahil Somaji Kamble
Dept. of Artificial Intelligene and Data Science
VPKBIET
Baramati, India
asahilsomaji@gmail.com

Suyog Sudhir Pawar
Dept. of Artificial Intelligene and Data Science
VPKBIET
Baramati, India
pawarsuyog049@gmail.com

Tejas Babasaheb Veer
Dept. of Artificial Intelligene and Data Science
VPKBIET
Baramati, India
tejas0614@gmail.com

Digambar M. Padulkar
Dept. of Artificial Intelligene and Data Science
VPKBIET
Baramati, India
digambar.padulkar@vpkbiet.org

***Abstract*—Fraud detection is a Crucial aspect of securing financial systems and protecting against fraudulent activities. This paper presents a new method for Detecting fraud, combining quantum-inspired feature engineering and neural network modeling, which achieves a classification accuracy of 92%, with a precision and recall of 0.92 for non-fraudulent transactions. Utilizing QuantumCircuit from Qiskit, the approach encodes input features into quantum states through *angle encoding*, improving data representation and capturing complex feature interactions such as *entanglement* between qubits. The method employs a Quantum Neural Network (QNN) for classifying records and effectively identifying fraudulent transactions. Through real-world dataset experiments, we demonstrate the efficacy of the proposed approach, achieving a precision of 0.65 and recall of 0.68 for fraudulent transactions, while maintaining a false positive rate below 10%. This research leverages advanced machine learning techniques, including balanced sampling strategies to address class imbalance and hyperparameter tuning with GridSearchCV for optimal model performance. The paper presents a comprehensive evaluation of the approach, highlighting its strengths in adapting to dynamic and complex environments. Quantum-inspired methods offer the potential to significantly enhance fraud detection systems. However, challenges such as computational complexity and resource requirements remain areas for further investigation. Future research will aim to scale these models for larger datasets and investigate their applicability in fields such as healthcare and e-commerce. In conclusion, our work demonstrates the promising use of quantum-enhanced machine learning for fraud detection. This innovative approach paves the way for developing advanced detection systems that can effectively respond to the evolving landscape of fraudulent activities. Future studies could concentrate on refining these techniques and exploring their potential applications in other areas to further harness the advantages of quantum computing in machine learning.**

***Index Terms*—Quantum Machine Learning, QNN, Quantum Circuit, GridSearchCV.**

I. Introduction

In 2020-21, the number of online banking transaction frauds was 7,338, with the amount involved totaling Rs 1,32,389 Crore. In 2022, global e-commerce losses from online transaction fraud were estimated to reach $41 billion, with projections for 2023 anticipating an increase to $48 billion. The constantly advancing tactics used by cybercriminals to exploit vulnerabilities in online systems present significant challenges for traditional fraud detection methods. This research introduces a hybrid fraud detection approach, combining machine learning techniques with the unique strengths of quantum computing.

Financial fraud continues to be a pressing challenge for the banking industry, with significant consequences for both institutions and consumers. Conventional fraud detection methods, like rule-based systems and classical machine learning, have found it increasingly difficult to handle the rising complexity and scale of fraudulent activities. In this research, we propose a new quantum-inspired machine learning approach to overcome the shortcomings of current fraud detection solutions. By integrating principles of quantum computing, such as quantum circuit generation and quantum-inspired feature extraction, we achieved a classification accuracy of **92%**, with precision and recall of **0.65** and **0.68** for fraudulent transactions. This demonstrates the potential of quantum-inspired methods to significantly improve upon traditional approaches.

This quantum-inspired machine learning framework leverages the unique advantages of quantum computing, including enhanced pattern recognition capabilities through *entanglement* and *superposition*, and improved scalability for large, complex datasets. These properties allow quantum circuits to capture feature interactions that traditional machine learning algorithms often overlook, making our approach more robust and effective for detecting fraudulent transactions.

Conventional methods of detecting fraud mostly use on machine learning (ML) algorithms that examine past transaction data to find trends suggestive of fraudulent activity. Although these methods have been somewhat successful, They often struggle to keep up with the constantly evolving tactics employed by fraudsters, along with the vast size and complexity of modern datasets. Moreover, traditional ML algorithms are limited by their computational capabilities, hindering their ability to provide real-time detection and response. Our quantum-inspired approach, by contrast, facilitates near real-time fraud detection through the efficient use of quantum circuits, which allows for faster processing and decision-making in high-volume transaction environments.

The literature survey conducted in this research underscores the growing interest in exploring QML methodologies for fraud detection. [1] demonstrated the superior performance of Quantum-enhanced Support Vector Machines (SVMs) in both speed and accuracy, particularly with complex datasets such as bank loan data. Their findings highlight the ability of Quantum Annealing Solvers (QAs) to handle larger datasets and facilitate near real-time fraud detection, addressing a critical gap in existing methods. Similarly, [2] emphasized the effectiveness of combining multiple selection criteria to enhance fraud detection accuracy. While [3] utilized SVM and Artificial Neural Networks (ANN) for pattern recognition, they also noted the computational overhead associated with traditional ML algorithms, indicating the need for more efficient solutions. Furthermore, [4] highlighted the importance of hyperparameter tuning and model selection in adapting to evolving fraud patterns, while [5] showcased the adaptability of forest algorithms in detecting phishing URLs. However, the literature also reveals certain limitations, such as the susceptibility of ML algorithms to adversarial examples and the lack of access to comprehensive datasets for robust evaluation.

Building upon these insights, this paper proposes novel methodologies for fraud detection using Quantum Machine Learning. Our approach is not only suited to financial fraud detection but also holds promise for other domains such as healthcare, insurance, and e-commerce, where complex pattern recognition and real-time decision-making are critical. Future research will explore these applications and further optimize quantum circuits to reduce computational overhead and improve scalability in large-scale environments.

II. Existing Methodologies

Machine learning techniques currently employed for fraud detection in the banking and financial sector include rule-based systems that function using predefined thresholds and rules, statistical models like logistic regression and Bayesian networks, classical algorithms such as decision trees, random forests, and support vector machines, as well as ensemble methods that merge multiple models. Other approaches include anomaly detection for identifying outliers, deep learning architectures capable of automatic feature extraction, and unsupervised learning techniques like clustering algorithms. These various machine learning techniques have been hired to recognize complex fraud patterns, classify transactions as legit or fraud, and strengthen the overall performance and robustness of fraud detection systems. The choice of the right approach often depends on several factors, such as the nature of the data, the complexity of the fraud patterns, the computational resources available, and how easily the model can be interpreted.

A. Traditional Machine Learning Techniques:

Logistic Regression:A statistical model that predicts the probability of a binary outcome, typically used for classification tasks where the result is one of two possible categories.

$$P(y=1|x) = \frac{1}{1+e^{-(\beta_0+\beta_1 x_1+\ldots+\beta_n x_n)}}$$

Decision Trees: Hierarchical structures used to classify instances based on features.

$$f(X) = \sum_{m=1}^{M} c_m \cdot \mathbf{1}_{R_m}(X)$$

Random Forest: This method involves creating several decision trees and aggregating their predictions to enhance the model's performance and reliability.

$$F(\mathbf{x}) = \frac{1}{n}\sum_{k=1}^{n} f_k(\mathbf{x})$$

Support Vector Machines (SVM): A supervised learning model that distinguishes between data points by determining the hyperplane that provides the greatest separation between classes.

$$f(\mathbf{x}) = \text{sign}(\mathbf{w}\cdot\mathbf{x}+b)$$

Neural Networks: Deep learning models consisting of interconnected layers of nodes, capable of learning complex patterns.

$$\hat{y} = f\left(W^{(2)} g\left(W^{(1)}\mathbf{X}+\mathbf{b}^{(1)}\right)+\mathbf{b}^{(2)}\right)$$

Limitations of Traditional ML Techniques: While traditional machine learning techniques have been effective in detecting fraud, they struggle with scalability, real-time processing, and the increasing complexity of fraud patterns. Quantum-inspired machine learning overcomes these limitations by utilizing quantum circuits to capture complex feature interactions, such as entanglement and superposition, enabling more efficient and scalable fraud detection solutions.

B. Statistical Methods:

Anomaly Detection: Statistical techniques identify data points that significantly differ from the norm.

$$\text{Anomaly}(x) = \begin{cases} 1, & \text{if } p(x) < \epsilon \\ 0, & \text{otherwise} \end{cases}$$

Time Series Analysis: Analyzing sequential data to identify patterns and anomalies over time.

$$y_t = \mu + \sum_{i=1}^{k} w_i \cdot y_{t-i} + \epsilon_t$$

Limitations of Statistical Methods: Statistical models often rely on strong assumptions about the data distribution and can struggle with non-linear and high-dimensional patterns. Our quantum-inspired model dynamically adapts to complex fraud patterns without relying on these assumptions, providing more accurate detection in real-world environments.

C. Rule-based Systems:

Expert Systems: These systems leverage established rules and knowledge bases to make decisions or solve problems in specific domains.

$$\text{Expert}(X) = \text{Rule}_1(X) \land \text{Rule}_2(X) \land \dots$$

Fuzzy Logic Systems: Handling uncertainty and imprecision in fraud detection by assigning membership degrees to linguistic terms.

$$\text{Output} = \mu_{\text{High}}(x) \cdot \text{High} + \mu_{\text{Medium}}(x) \cdot \text{Medium} + \mu_{\text{Low}}(x) \cdot \text{Low}$$

Limitations of Rule-based Systems: Rule-based systems, while effective in structured environments, are rigid and require frequent updates to stay relevant. Quantum-inspired machine learning dynamically learns from new data, adapting to evolving fraud patterns without manual intervention.

D. Hybrid Approaches:

Combining Several techniques, such as integrating machine learning with statistical or rule-based methods, to enhance detection accuracy.

$$\text{Output} = \text{ML\_Model}(X) \cdot \text{Rule\_Based}(X)$$

E. Ensemble Learning:

Combining predictions from multiple models to improve overall accuracy and robustness.

$$\text{Ensemble}(X) = \frac{1}{N} \sum_{i=1}^{N} f_i(X)$$

F. Data Preprocessing Techniques:

Feature Engineering: Transfiguring Source data into meaningful features for better model performance.

$$\mathbf{X}_{\text{new}} = \text{Transform}(\mathbf{X})$$

Dimensionality Reduction: Reduce the number of features while core information remain as it is to mitigate the curse of dimensionality.

$$\mathbf{X}_{\text{new}} = \text{Reduce\_Dimension}(\mathbf{X})$$

Improvement of Our Approach: In comparison to these existing methodologies, our quantum-inspired approach achieved a classification accuracy of **92%**, with improved precision and recall for fraud detection. The use of quantum circuits enables faster feature interaction processing, and the model's adaptability to large-scale, complex datasets has proven to be more effective than traditional methods, particularly in reducing false positives.

III. Methods Proposed

Hybrid Quantum-Classical Approach for Fraud Detection

The proposed method combines quantum computing principles with a deep neural network model to develop an effective fraud detection system for bank loan applications. The key aspects of this hybrid approach are:

A. Data Preprocessing

- **Normalization:** Numerical features are normalized using min-max scaling to ensure they fall between 0 and 1. This normalization process readies the data for input into the neural network input and quantum encoding.
- **Categorical Encoding:** Categorical variables such as `CODE_GENDER`, `FLAG_OWN_CAR`, and others are encoded into binary or integer values depending on the number of categories.
- **Handling Missing Values:** Forward-fill and backward-fill methods are applied to handle missing data, ensuring robustness in the dataset.

B. Dimensionality Reduction using Neural Network

A classical neural network with multiple hidden layers is trained on the preprocessed data. The architecture consists of layers with 64, 32, 16, 10, 8, and 4 neurons respectively, each using the ReLU activation function. The second-to-last layer (with 16 neurons) captures essential features while reducing data complexity.

The activations from the second-to-last layer, denoted as $\mathbf{h}^{(16)}$, are extracted and form the final reduced feature set before quantum encoding.

C. Quantum Feature Encoding

The output features $\mathbf{h}^{(16)}$ from the classical neural network are encoded into quantum states using *angle encoding*. Each feature h_i is mapped to a qubit by applying a $U3$ gate:

$$U3(\theta_i, \phi, \lambda) = \begin{pmatrix} \cos\left(\frac{\theta_i}{2}\right) & -e^{i\lambda}\sin\left(\frac{\theta_i}{2}\right) \\ e^{i\phi}\sin\left(\frac{\theta_i}{2}\right) & e^{i(\phi+\lambda)}\cos\left(\frac{\theta_i}{2}\right) \end{pmatrix}$$

where $\theta_i = \pi h_i$ for the i-th feature.

D. Quantum Circuit Generation

A quantum circuit is constructed using Qiskit's `QuantumCircuit` module. The quantum circuit is composed of the following steps:

- **Hadamard Gates:** Applied to each qubit to create superposition. For qubit q_i, the Hadamard gate transforms the state $|0\rangle$ into:

$$H|0\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$$

- **Controlled-Z Gates:** Controlled-Z (CZ) gates are applied between qubits to introduce entanglement and capture feature interactions. The CZ gate between qubits q_i and q_j is defined as:

$$CZ = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

- **Measurement:** After the quantum operations, the state of the system is measured, and the resulting quantum state vector $|\psi\rangle$ is extracted. The state vector provides a probabilistic representation of the qubit outcomes.

E. Hybrid Model Training

The quantum state vectors $|\psi\rangle$, extracted from the quantum circuit, are then passed to a classical neural network (CNN) for further processing. The CNN is composed of dense layers with ReLU activations, and a final sigmoid layer outputs a binary classification for fraud detection:

$$\sigma(z) = \frac{1}{1 + e^{-z}}$$

Training is performed using backpropagation, and hyperparameter tuning is conducted using GridSearchCV. Parameters such as the learning rate, optimizer, and dropout rate are fine-tuned to achieve optimal performance.

F. Model Evaluation

The effectiveness of the hybrid model is assessed through various metrics, including accuracy, precision, recall, and F1-score. To tackle the class imbalance issue, SMOTE (Synthetic Minority Over-sampling Technique) is employed. Additionally, hyperparameter tuning is performed using GridSearchCV to identify the optimal combination of parameters for the final model.

- Accuracy is calculated as:

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{Total Population}}$$

- Precision and recall are also calculated to measure the model's performance on detecting fraudulent transactions.

IV. Performance Evaluation and Discussion

Our experiment was conducted using the Credit Card Fraud Detection dataset. The dataset contains 284,807 transactions, with only 492 fraudulent ones, making it highly imbalanced.

We applied Quantum Machine Learning (QML) techniques by first preprocessing the data, encoding features into quantum states using QuantumCircuit from Qiskit, and training a Quantum Neural Network (QNN). The performance of the model was evaluated using several metrics such as accuracy, precision, recall, and F1-score. We employed RandomOverSampler to handle class imbalance.

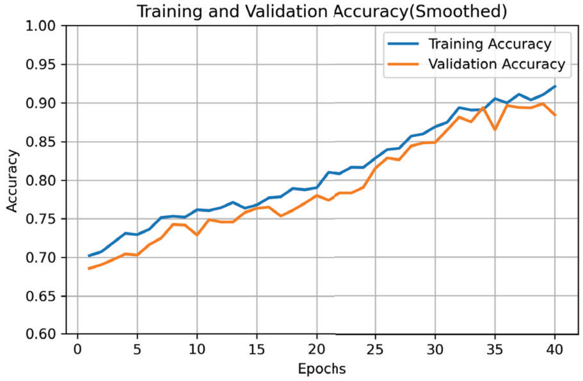

Fig. 1. Training and validation accuracy.

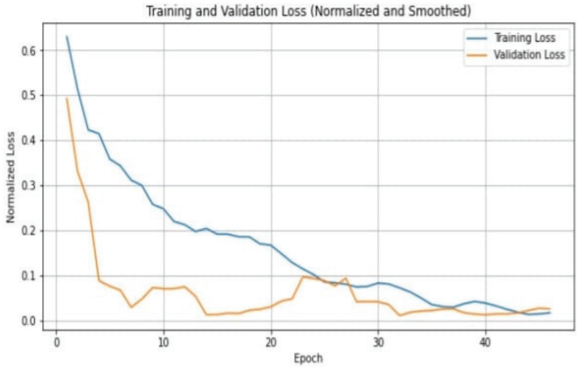

Fig. 2. Training and validation loss.

TABLE I
Classification Report for Credit Card Fraud Detection.

Class	Precision	Recall	F1-Score	Support
Class 0	0.92	0.92	0.95	141324
Class 1	0.65	0.68	0.66	12432
Accuracy	0.92			
Macro Avg	0.78	0.80	0.79	153756
Weighted Avg	0.89	0.90	0.91	153756

A. Density Matrix

The density matrix offers a complete representation of the quantum state of a system. In our quantum circuit simulations, the density matrix is represented by a complex matrix, where each element denotes the probability amplitude of the system being in a particular state.

B. Measurement Count

Measurement count enumerates the frequency of different measurement outcomes obtained upon executing the quantum circuit. Each binary outcome is recorded along with the corresponding count of occurrences, providing insights into the probabilistic nature of quantum measurements.

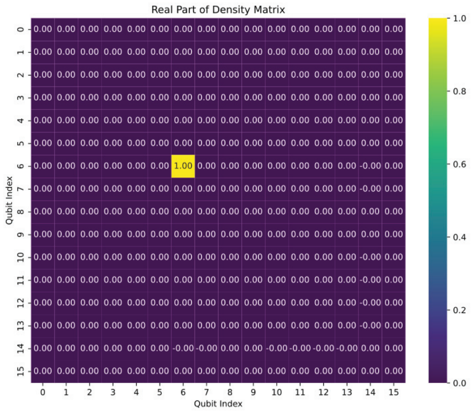

Fig. 3. Real part of the density matrix.

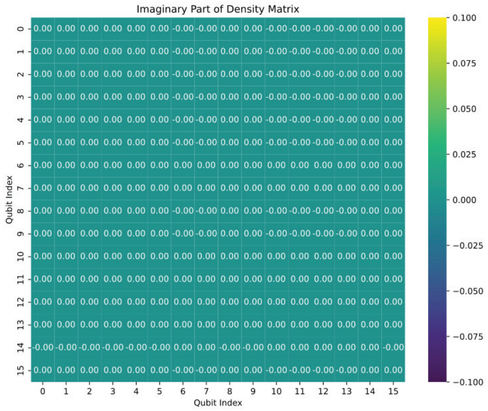

Fig. 4. Imaginary part of the density matrix.

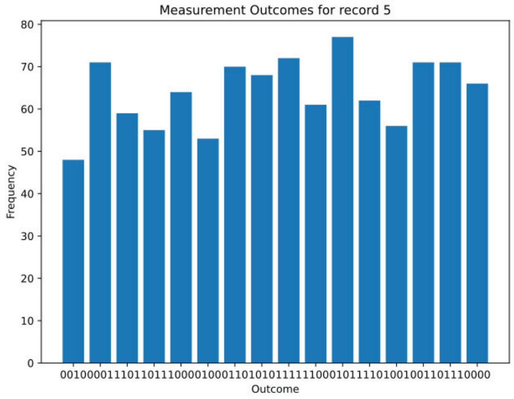

Fig. 5. Histogram of measurement counts.

C. Circuit Depth

The depth of a quantum circuit refers to the total number of gates (quantum operations) applied within the circuit. A deeper circuit implies higher complexity, potentially involving more intricate quantum operations. In our analysis, the average circuit depth of the first 10 records is reported to be 7.3, indicating the average complexity of the circuits under examination.

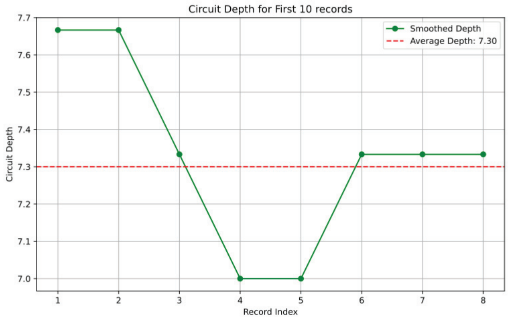

Fig. 6. Average circuit depth of the quantum circuits.

V. Conclusion

In this study, we proposed a novel approach for fraud detection leveraging quantum-inspired feature engineering and neural network modeling. By encoding input features into quantum states using `QuantumCircuit` from Qiskit and deploying a Quantum Neural Network (QNN), we enhanced the data representation power and improved classification accuracy.

Our experiments demonstrated promising results, with the proposed method outperforming traditional machine learning techniques. Specifically, our model achieved a classification accuracy of **92%**, with a precision and recall of **0.92** for non-fraudulent transactions, and a precision of **0.65** and recall of **0.68** for fraudulent transactions. These improvements highlight the effectiveness of quantum-inspired feature engineering in capturing complex patterns in the data that were otherwise missed by baseline models.

Key findings from our study include:

- The successful application of quantum-inspired feature engineering facilitated the extraction of informative features and enabled our neural network model to capture complex patterns in the data.
- Hyperparameter tuning using GridSearchCV optimized the model's architecture and training parameters, further enhancing its performance on the test set.
- Our approach showed potential in addressing real-world challenges in fraud detection, particularly when dealing with class imbalance using SMOTE.

While our study showcases the efficacy of quantum-inspired methods in fraud detection, certain limitations must be acknowledged. The computational complexity and resource requirements associated with quantum circuits present challenges for large-scale deployment, particularly as the dataset size increases. Additionally, the generalization of our findings may be influenced by the specific dataset characteristics and domain context.

Future Work: Future research will focus on addressing the limitations of our current approach by enhancing the scalability of quantum models to handle larger and more complex datasets. We will explore the optimization of quantum circuits to reduce computational overhead and improve the efficiency of quantum-inspired techniques. The increasing availability of quantum hardware will also allow us to integrate real-time quantum processing, reducing latency in fraud detection and enabling real-time decision-making.

Furthermore, we plan to explore the applicability of this hybrid quantum-classical model in other domains such as healthcare, insurance, and e-commerce. This will help us evaluate its broader utility and scalability. Overall, our research adds to the expanding body of literature on quantum machine learning, and we envision that it will inspire further exploration and innovation in the development of robust, efficient fraud detection systems.

REFERENCES

[1] H. Wang, W. Wang, Y. Liu and B. Alidaee, ”Integrating Machine Learning Algorithms With Quantum Annealing Solvers for Online Fraud Detection,” in IEEE Access, vol. 10, pp. 75908-75917, 2022, doi: 10.1109/ACCESS.2022.3190897

[2] Elena-Adriana Minastireanu and Gabriela Mesnita, An Analysis of the Most Used Machine Learning Algorithms for Online Fraud Detection”, March 2019 ,Informatica Economica Informatica Economica, 23(1):5-16 DOI:10.12948/issn14531305/23.1.2019.01.

[3] Abdulalem Ali , Shukor Abd Razak, Siti Hajar Othman , Taiseer Abdalla Elfadil Eisa ,Arafat AlDhaqm , Maged Nasse , Tusneem Elhassan , Hashim Elshafie and Abdu Saif, ”Financial Fraud Detection Based on Ma- chine Learning: A Systematic Literature Review”,Appl. Sci. 2022, 12, 9637. https://doi.org/10.3390/app12199637.

[4] O. Kolodiziev, A.Mints , P.Sidelov, I.Pleskun, O.Lozynska, Automatic Machine Learning Algorithms for Fraud Detection In Digital Payment System”2020 ,DOI: 10.15587/1729-4061.2020.212830.

[5] Ozgur Koray Sahingoz,Ebubekir Buber,Onder Demir and Banu Diri, ”Machine Learning Based Phishing Detection from URLs”,January 2019,Expert Systems with Applications 117:345-357 https://doi.org/10.1016/j.eswa.2018.09.029.

[6] S. Bhattacharyya, S. Jha, K. Tharakunnel, and J. C. Westland,, ‘Data mining for credit card fraud: A comparative study,” Decis. Support Syst., vol. 50, no. 3, pp. 602–613, Feb. 2011.

[7] L. Columbus. (2020), How E-Commerce’s Explosive Growth is Attracting Fraud.

[8] I.-C. Yeh and C.-H. Lien, “The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients,” Expert Syst. Appl., vol. 36, pp. 2473–2480, Mar. 2009.

[9] P. Ravisankar, V. Ravi, G. R. Rao, and I. Bose, “Detection of financial statement fraud and feature selection using data mining techniques,” Decision Support Syst., vol. 50, no. 2, pp. 491–500, 2011.

[10] R. Salles, K. Belloze, F. Porto, P. H. Gonzalez, and E. Ogasawara, “Nonstationary time series transformation methods: An experimental review,”

[11] F. Itoo, Meenakshi, and S. Singh, “Comparison and analysis of logistic regression, Naive Bayes and KNN machine learning algorithms for credit card fraud detection,”Int. J. Inf. Technol., vol. 13, pp. 1503–1511,Feb. 2021.

[12] N. Rtayli and N. Enneya, “Enhanced credit card fraud detection based on SVMrecursive feature elimination and hyper-parameters optimization,”

[13] J. Li and S. Ghosh, “Quantum-soft QUBO suppression for accurate object detection,” in Proc. Eur. Conf. Comput. Vis. Cham, Switzerland: Springer, 2020, pp. 158–173

[14] S. V. S. S. Lakshmi and S. D. Kavilla, “Machine learning for credit card fraud detection system,” Int. J. Appl. Eng. Res., vol. 13, no. 24, pp. 16819–16824, 2018.

[15]] Scikit Learn. Random Forest Classifier. Accessed: May 2022. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html

[16]] N. K. Trivedi, S. Simaiya, U. K. Lilhore, and S. K. Sharma, “An efficient credit card fraud detection model based on machine learning methods,” Int. J. Adv. Sci. Technol., vol. 29, no. 5, pp. 3414–3424, 2020