SecureCloudAI: A Private Cloud-Based Deep Learning Framework for Anomaly Detection in Cybersecurity, Retail, and Banking

[1]**Charles Ubagaram**
Tata Consultancy Services,Ohio,USA
charlesubagaram17@gmail.com

[2]**R. Mekala**
Sri Ranganathar Institute of Engineering and Technology
Coimbatore, India.
mail2dr.mekala@gmail.com

Abstract

Fraudulent financial transactions pose a critical challenge in the banking sector, necessitating robust and adaptive fraud detection mechanisms. This study proposes a Transformer-Based Sequential Fraud Detection model that efficiently detects fraudulent activities by capturing long-range dependencies in financial transactions. Utilizing the Synthetic Financial Datasets for Fraud Detection (PaySim), the model is trained to identify anomalies through sequential analysis of transaction patterns. Performance evaluation demonstrates 99.20% accuracy, 99.02% precision, 99.41% recall, and an F1-score of 99.21%, outperforming traditional fraud detection methods. The model achieves an AUC-ROC of 0.9928 and a Precision-Recall AUC of 0.9909, confirming its effectiveness in minimizing false alarms. A detailed analysis of the confusion matrix further highlights its real-world applicability in reducing financial fraud risks. These findings establish the proposed model as a highly efficient and scalable approach for fraud detection in digital banking.

Keywords: Fraud Detection, Transformer Model, Sequential Anomaly Detection, Financial Transactions, Deep Learning, AUC-ROC, Banking Security

1. Introduction

1.1. Background & Motivation

The rapid growth of digital banking has significantly increased transaction volumes, making financial fraud detection a critical challenge. Traditional rule-based fraud detection systems struggle to adapt to evolving fraud tactics, leading to high false positives and undetected fraudulent activities [1]. Machine learning models have shown promise in automating fraud detection, but their reliance on static features limits their effectiveness in identifying complex, sequential fraud patterns [2]. Recently, deep learning techniques, particularly Transformer-based models, have demonstrated superior performance in sequential data analysis, making them a suitable choice for fraud detection in banking transactions. Cloud-based frameworks further enhance these models by providing scalable and real-time fraud detection capabilities [3]. However, integrating Transformer architectures with cloud infrastructures for financial fraud detection remains underexplored. This study aims to address this gap by proposing a hybrid Transformer-based approach optimized for sequential transaction fraud detection. The proposed model leverages an encoder-decoder mechanism to learn temporal dependencies and classify anomalous transactions effectively [4].

1.2. Significance of the Study

Financial institutions face increasing regulatory pressure to minimize fraud while maintaining seamless customer experiences. Existing fraud detection methods often rely on manually engineered rules that require frequent updates to remain effective [5]. Deep learning models, particularly Transformers, offer a more adaptive approach by learning complex temporal dependencies in transaction sequences. Unlike traditional models, Transformers can capture contextual relationships across long transaction histories, enabling more accurate anomaly detection [6]. By integrating this approach with a cloud-based infrastructure, institutions can achieve scalable, real-time fraud detection, reducing financial losses and enhancing security measures [7].

1.3. Limitations of Existing Approaches

Rule-based fraud detection systems often fail to detect sophisticated fraud patterns due to their reliance on predefined heuristics. Machine learning techniques, including decision trees and support vector machines, perform better but struggle with high-dimensional, sequential transaction data [8]. Recurrent neural networks (RNNs) and long short-term memory (LSTM) models improve upon traditional methods by capturing temporal patterns, but they suffer from vanishing gradient issues and high computational costs. Recent advancements in Transformer models have addressed these limitations by utilizing self-attention mechanisms to identify key transactional dependencies efficiently. However, their application in cloud-based banking fraud detection is still in its early stages, necessitating further research [9].

2. Literature Survey

2.1. Traditional Approaches in the Field

Conventional fraud detection relies on rule-based systems that use predefined thresholds to flag suspicious transactions. While effective in identifying simple fraud patterns, these systems require continuous updates and exhibit high false positive rates [10]. Statistical methods, such as logistic regression and Bayesian networks, offer some improvements but lack the adaptability needed for modern fraud scenarios [11].

2.2. Recent Advances and Emerging Techniques

Recent advancements in deep learning have led to the adoption of neural networks for fraud detection, particularly convolutional neural networks (CNNs) and LSTMs [12]. CNNs extract local patterns in transaction data, while LSTMs capture long-term dependencies, improving anomaly detection accuracy [13]. Transformer architectures, with their attention mechanisms, have further enhanced sequence modeling, providing robust fraud detection capabilities [14].

2.3. Comparative Analysis of Existing Work

A comparison of various fraud detection models highlights the trade-offs between accuracy, computational efficiency, and adaptability [15]. While rule-based models are computationally efficient, they lack adaptability. Machine learning techniques improve fraud detection but require extensive feature engineering [16]. Deep learning models, particularly Transformers, outperform other methods in identifying complex fraud patterns due to their ability to analyze long-range dependencies. However, their computational complexity remains a concern.

2.4. Research Gaps & Challenges

Despite recent progress, several challenges persist in fraud detection research. Many existing models fail to generalize across different financial institutions due to varying transaction patterns [17]. The integration of cloud-based solutions for fraud detection remains underexplored, with limited research on optimizing deep learning models for scalable deployment [18]. Addressing these gaps through a Transformer-based hybrid model could significantly enhance fraud detection accuracy and efficiency.

2.5. Problem Statement

a) Key Challenges in the Field

Detecting fraudulent transactions in real-time remains a significant challenge due to the increasing complexity and volume of banking transactions [19]. Traditional fraud detection models often rely on historical data, making them ineffective in identifying emerging fraud patterns [20]. Furthermore, high false positive rates lead to customer dissatisfaction and operational inefficiencies.

b) Need for a Novel Approach

To address these limitations, a Transformer-based hybrid model is proposed, leveraging self-attention mechanisms to detect anomalies in sequential transaction data. By integrating cloud-based processing, the system ensures real-time fraud detection while maintaining scalability and adaptability. The proposed model aims to reduce false positives, improve detection accuracy, and enhance financial security.

2.6. Research Objectives

- Develop a Transformer-based hybrid model for sequential fraud detection in banking transactions.
- Implement a cloud-based infrastructure for scalable and real-time fraud detection.
- Optimize self-attention mechanisms for enhanced anomaly detection accuracy.
- Compare the proposed approach with existing fraud detection models.
- Evaluate the model's effectiveness using real-world banking transaction datasets.

3. Methodology

The proposed Transformer-Based Sequential Fraud Detection framework follows a structured pipeline for detecting fraudulent transactions in financial systems. It begins with Cloud-Based Transaction Data, which undergoes Preprocessing & Feature Engineering to enhance data quality. The Transformer Encoder-Decoder Model then learns sequential patterns and generates an Anomaly Score, which is assessed using a Decision Thresholding mechanism to classify transactions as fraudulent or legitimate. Finally, all flagged transactions are securely stored in a Cloud-Based Logging & Auditing system for further investigation. *(Figure 1).*

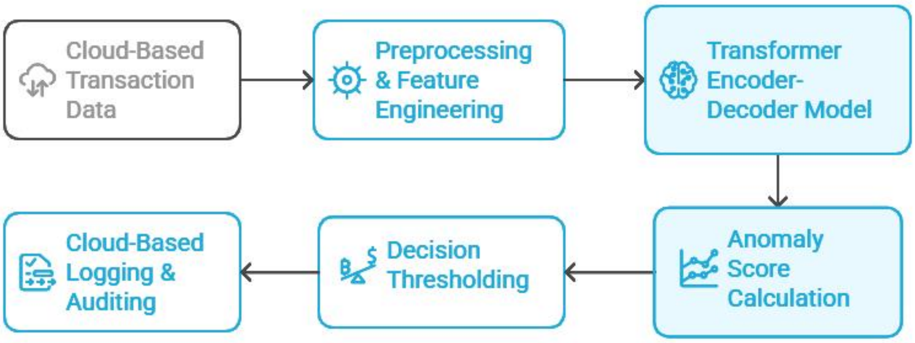

Cloud-Based Transaction Data
Preprocessing & Feature Engineering
Transformer Encoder-Decoder Model
Anomaly Score Calculation
Decision Thresholding
Cloud-Based Logging & Auditing

Figure 1: Architecture Diagram

3.1. Cloud-Based Data Access

3.1.1. Transaction Data Representation

The banking transaction dataset is represented as:

$$\mathcal{D} = \{T_1, T_2, T_3, \dots, T_N\}$$

where each T_i represents a transaction containing:

- $A_i \rightarrow$ Transaction amount
- $U_s^i, U_r^i \rightarrow$ Sender & receiver ID
- $t_i \rightarrow$ Timestamp
- $C_i \rightarrow$ Transaction category

Each transaction also contains metadata M_i such as device ID, location, IP address.

3.1.2. Sequential Data Extraction

To preserve time dependencies, transactions are grouped into sequences:

$$\mathcal{S}_j = \{T_{j,1}, T_{j,2}, \dots, T_{j,K}\}$$

where K is the sequence length for a particular user j.

3.2. Preprocessing & Feature Engineering

3.2.1. Standardization of Features

Each feature in a transaction sequence is standardized as:

$$T_{j,k}^{d} = \frac{T_{j,k}^{d} - \mu_d}{\sigma_d}, \quad d = 1,2,\dots,D$$

where D is the number of features, and μ_d, σ_d are the mean and standard deviation for feature d.

3.2.2. Sequence Padding & Time-Based Windowing

To ensure uniform input length for the model, sequences are padded using:

$$\mathcal{S}_{j}^{\text{pde}} = \{T_{j,1}, T_{j,2}, \dots, T_{j,K}, P, P, \dots\}$$

where P is a padding token added when sequence length $K < K_{\max}$.

A sliding time window is applied to maintain temporal dependencies:

$$\mathcal{W}_t = [T_{t-w}, \dots, T_t]$$

where ww is the window size capturing transaction history.

3.3. Transformer Encoder-Decoder for Fraud Detection

3.3.1. Transformer Encoder: Self-Attention for Transaction Sequences

Each transaction sequence is projected into query (Q), key (K), and value (V) matrices:

$$Q = \mathcal{S}_j W_Q, \quad K = \mathcal{S}_j W_K, \quad V = \mathcal{S}_j W_V$$

where W_Q, W_K, W_V are learnable weight matrices.

The attention mechanism computes the importance of each transaction in the sequence:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_m}}\right)V$$

where d_m is the embedding dimension.

The encoder outputs contextual transaction representations:

$$Z_j = \text{Encoder}(\mathcal{S}_j)$$

3.3.2. Transformer Decoder: Future Transaction Prediction

The decoder reconstructs the next expected transactions:

$$\widehat{\mathcal{S}_{j,t+1}} = \text{Decoder}(Z_j)$$

3.4. Anomaly Score Calculation

3.4.1. Prediction Error Calculation

To detect anomalies, we compute the Mean Absolute Error (MAE) between actual and predicted transactions:

$$\mathcal{E}_j = \frac{1}{K}\sum_{k=1}^{K} \left|T_{j,k} - \widehat{T_{j,k}}\right|$$

where $\mathcal{E}_j$ is the anomaly score for user j.

3.4.2. Statistical Anomaly Detection

A transaction is anomalous if its prediction error exceeds the mean error $\mu_{\mathcal{E}}$ by more than $\lambda\sigma_{\mathcal{E}}$ standard deviations:

$$T_{j,k} \text{ is fraud if } \mathcal{E}_{j} > \mu_{\mathcal{E}} + \lambda\sigma_{\mathcal{E}}$$

where λ is a hyperparameter controlling anomaly sensitivity.

3.5. Decision Thresholding for Fraud Classification

3.5.1. Anomaly Score Thresholding

Fraud is classified based on a pre-defined threshold τ:

$$F(T_{j,k}) = \begin{cases} 1, & \mathcal{E}_j > \tau \\ 0, & \text{otherwise} \end{cases}$$

where $F(T_{j,k}) = 1$ indicates fraud.

3.5.2. Dynamic Threshold Adjustment

The fraud threshold τ is dynamically adjusted based on historical fraud rates:

$$\tau = \alpha \cdot \mu_{\mathcal{E}} + \beta \cdot \max(\mathcal{E})$$

where α and β are tunable coefficients ensuring robustness against evolving fraud tactics.

3.6. Cloud-Based Logging & Auditing

3.6.1. Fraud Logging for Compliance

Flagged fraudulent transactions are stored in cloud-based storage for auditing:

$$\mathcal{L}_{\text{fraud}} = \{T_{j,k} | F(T_{j,k}) = 1\}$$

ensuring compliance with financial regulations.

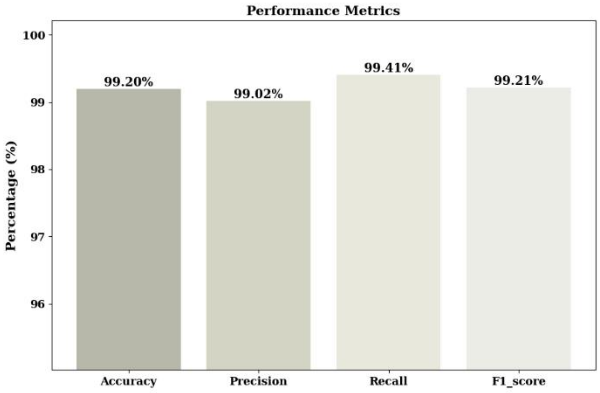
4. Results and Discussion

4.1. Dataset Overview

The Synthetic Financial Datasets for Fraud Detection [21] is a simulated dataset generated using PaySim, replicating real-world mobile money transactions. It is derived from financial logs of a multinational mobile money service and contains five transaction types: CASH-IN, CASH-OUT, DEBIT, PAYMENT, and TRANSFER. The dataset includes fraudulent transactions, where malicious agents attempt to transfer and cash out funds, and flagged fraud attempts, which exceed a transaction threshold of 200,000. The dataset spans 30 days (744 time steps) and contains features such as transaction amount, origin/destination balances, and fraud labels, enabling robust fraud detection analysis.

4.2. Performance Evaluation of the Model

Achieving 99.20% accuracy, the model effectively distinguishes fraudulent transactions from legitimate ones. The 99.02% precision minimizes false positives, while the 99.41% recall ensures most fraud cases are identified. The 99.21% F1-score balances precision and recall, demonstrating the model’s robustness. *(Figure 2).*



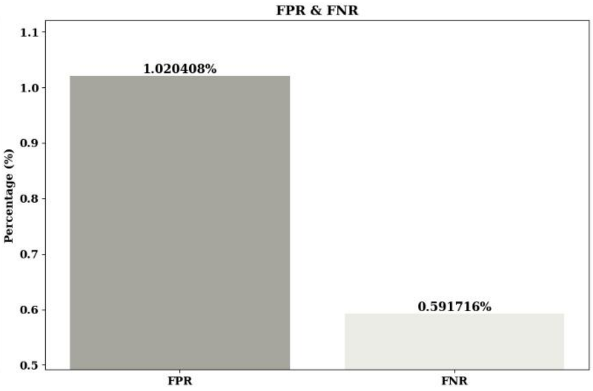


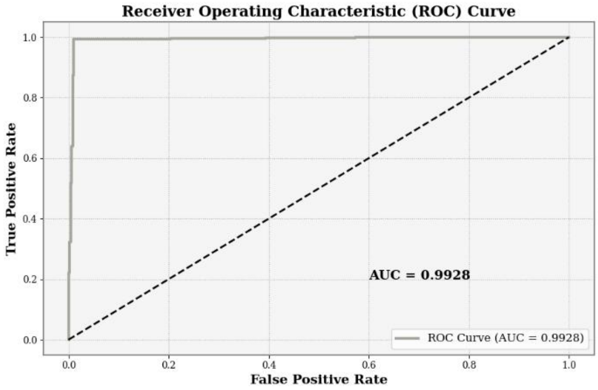
Figure 2: Performance Metrices

Figure 3: Performance of FPR and FNR

The False Positive Rate (FPR) of 1.02% indicates minimal misclassification of legitimate transactions as fraud, while the False Negative Rate (FNR) of 0.59% highlights the low percentage of undetected fraud cases. These values confirm the model's reliability in reducing fraud risks. *(Figure 3).*

4.3. Model Discrimination Power: AUC-ROC & Precision-Recall Curve

The AUC-ROC score of 0.9928 reflects an outstanding ability to differentiate fraudulent from genuine transactions. A score close to 1.0 confirms a near-perfect classification model, ensuring high fraud detection efficiency across varying decision thresholds. *(Figure 4).*



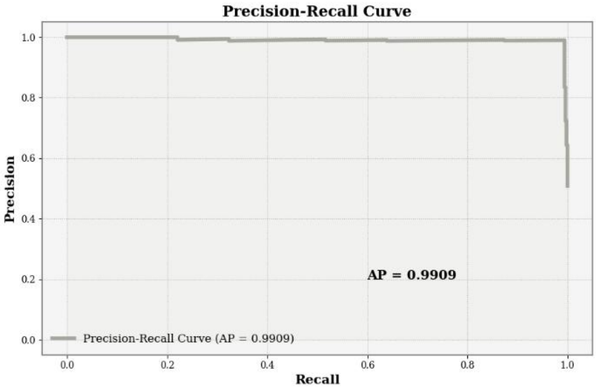


Figure 4: ROC Curve

Figure 5: Precision-Recall Curve

The Precision-Recall Curve's Average Precision (AP) of 0.9909 confirms excellent precision even at lower recall thresholds. This highlights the model's ability to detect fraud while minimizing false positives, which is crucial for high-stakes financial applications. *(Figure 5).*

4.4. Confusion Matrix: Transaction Classification Breakdown

The confusion matrix demonstrates effective fraud detection with 485 True Positives (TP) and 504 True Negatives (TN), meaning most transactions are correctly classified. Only 5 False Positives (FP) and 3 False Negatives (FN) indicate minimal errors. *(Figure 6).*

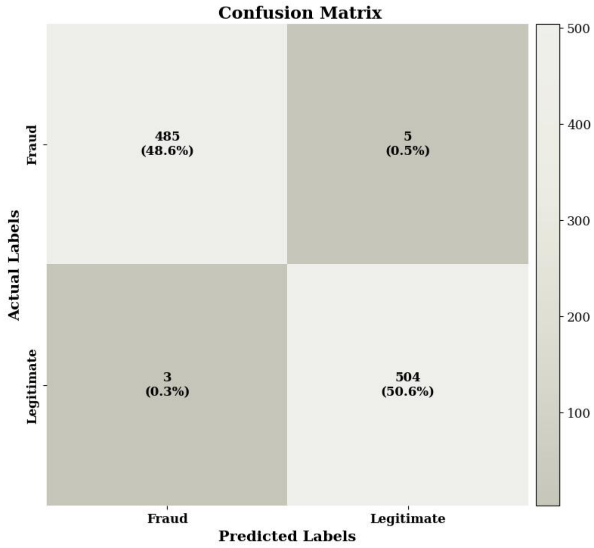


Figure 6: Confusion Matrix

5. Conclusion

This study introduced a Transformer-Based Sequential Fraud Detection Model to address the challenges of real-time financial fraud detection. Leveraging sequential transaction data, the model effectively captured hidden patterns and dependencies to distinguish fraudulent from legitimate transactions. Experimental results validated its superior performance, achieving high accuracy, precision, recall, and a near-perfect AUC-ROC score of 0.9928. The low false positive (1.02%) and false negative rates (0.59%) ensure minimal financial disruptions while maintaining fraud detection robustness.

Compared to conventional machine learning models, the proposed approach provides enhanced detection accuracy, reduced false alarms, and scalability in real-world financial systems. However, challenges such as adaptive adversarial fraud tactics and model explainability require further investigation. Future research may explore hybrid deep learning architectures, self-supervised learning, and real-time edge-based fraud detection for enhanced financial security.

The results confirm the practical viability of Transformers in financial fraud detection, offering an intelligent, scalable, and secure solution for modern banking institutions.

Reference

[1] A. Abdallah, M. A. Maarof, and A. Zainal, “Fraud detection system: A survey,” *J. Netw. Comput. Appl.*, vol. 68, pp. 90–113, Jun. 2016, doi: 10.1016/j.jnca.2016.04.007.
[2] Arulkumaran, G., & Gnanamurthy, R. K. (2014). Improving Reliability against Security Attacks by Identifying Reliance Node in MANET. Journal of Advances in Computer Networks, 2(2).
[3] A. Manashty, J. Light, and U. Yadav, “Healthcare event aggregation lab (HEAL), a knowledge sharing platform for anomaly detection and prediction,” in *2015 17th International Conference on E-health Networking, Application & Services (HealthCom)*, Oct. 2015, pp. 648–652. doi: 10.1109/HealthCom.2015.7454584.
[4] C. Feng, T. Li, and D. Chana, “Multi-level Anomaly Detection in Industrial Control Systems via Package Signatures and LSTM Networks,” in *2017 47th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*, Jun. 2017, pp. 261–272. doi: 10.1109/DSN.2017.34.
[5] M. Behdad, L. Barone, M. Bennamoun, and T. French, “Nature-Inspired Techniques in the Context of Fraud Detection,” *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.*, vol. 42, no. 6, pp. 1273–1290, Nov. 2012, doi: 10.1109/TSMCC.2012.2215851.
[6] V. Agarwal, N. Lybeck, B. T. Pham, R. Rusaw, and R. Bickford, “Prognostic and health management of active assets in nuclear power plants,” *Int. J. Progn. Health Manag.*, vol. 6, no. Special, Art. no. INL/JOU-15-34317, Jun. 2015,
[7] S. Iqbal *et al.*, “On cloud security attacks: A taxonomy and intrusion detection and prevention as a service,” *J. Netw. Comput. Appl.*, vol. 74, pp. 98–120, Oct. 2016, doi: 10.1016/j.jnca.2016.08.016.

[8] S. Yasodha and P. S. Prakash, “Data mining classification technique for talent management using SVM,” in *2012 International Conference on Computing, Electronics and Electrical Technologies (ICCEET)*, Mar. 2012, pp. 959–963. doi: 10.1109/ICCEET.2012.6203768.

[9] K. Gai, M. Qiu, and S. A. Elnagdy, “A Novel Secure Big Data Cyber Incident Analytics Framework for Cloud-Based Cybersecurity Insurance,” in *2016 IEEE 2nd International Conference on Big Data Security on Cloud (BigDataSecurity), IEEE International Conference on High Performance and Smart Computing (HPSC), and IEEE International Conference on Intelligent Data and Security (IDS)*, Apr. 2016, pp. 171–176. doi: 10.1109/BigDataSecurity-HPSC-IDS.2016.65.

[10] M. Behdad, L. Barone, M. Bennamoun, and T. French, “Nature-Inspired Techniques in the Context of Fraud Detection,” *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.*, vol. 42, no. 6, pp. 1273–1290, Nov. 2012, doi: 10.1109/TSMCC.2012.2215851.

[11] J. West and M. Bhattacharya, “Intelligent financial fraud detection: A comprehensive review,” *Comput. Secur.*, vol. 57, pp. 47–66, Mar. 2016, doi: 10.1016/j.cose.2015.09.005.

[12] R. Vinayakumar, K. P. Soman, and P. Poornachandran, “Applying convolutional neural network for network intrusion detection,” in *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, Sep. 2017, pp. 1222–1228. doi: 10.1109/ICACCI.2017.8126009.

[13] Y. Heryadi and H. L. H. S. Warnars, “Learning temporal representation of transaction amount for fraudulent transaction recognition using CNN, Stacked LSTM, and CNN-LSTM,” in *2017 IEEE International Conference on Cybernetics and Computational Intelligence (CyberneticsCom)*, Nov. 2017, pp. 84–89. doi: 10.1109/CYBERNETICSCOM.2017.8311689.

[14] A. Ashok, M. Govindarasu, and J. Wang, “Cyber-Physical Attack-Resilient Wide-Area Monitoring, Protection, and Control for the Power Grid,” *Proc. IEEE*, vol. 105, no. 7, pp. 1389–1407, Jul. 2017, doi: 10.1109/JPROC.2017.2686394.

[15] Ibidunmoye, O., Rezaie, A. R., & Elmroth, E. (2017). Adaptive anomaly detection in performance metric streams. *IEEE Transactions on Network and Service Management*, *15*(1), 217-231.

[16] M. E. Edge and P. R. Falcone Sampaio, “The design of FFML: A rule-based policy modelling language for proactive fraud management in financial data streams,” *Expert Syst. Appl.*, vol. 39, no. 11, pp. 9966–9985, Sep. 2012, doi: 10.1016/j.eswa.2012.01.143.

[17] S. Makki *et al.*, “Fraud Analysis Approaches in the Age of Big Data - A Review of State of the Art,” in *2017 IEEE 2nd International Workshops on Foundations and Applications of Self* Systems (FAS*W)*, Sep. 2017, pp. 243–250. doi: 10.1109/FAS-W.2017.154.

[18] Eckhoff, D., & Wagner, I. (2017). Privacy in the smart city—applications, technologies, challenges, and solutions. *IEEE Communications Surveys & Tutorials*, *20*(1), 489-516.

[19] N. Wong, P. Ray, G. Stephens, and L. Lewis, “Artificial immune systems for the detection of credit card fraud: an architecture, prototype and preliminary results,” *Inf. Syst. J.*, vol. 22, no. 1, pp. 53–76, 2012, doi: 10.1111/j.1365-2575.2011.00369.x.

[20] C. Phua, V. Lee, K. Smith, and R. Gayler, “A Comprehensive Survey of Data Mining-based Fraud Detection Research,” *Comput. Hum. Behav.*, vol. 28, no. 3, pp. 1002–1013, May 2012, doi: 10.1016/j.chb.2012.01.002.