A Hybrid Approach for Fraud Detection in Digital Wallet Transactions Using Adversarial Autoencoders and Gated Recurrent Units

Shaik Janbhasha
Department of Computer Science and Engineering (Data Science), CVR College of Engineering, Telangana, India
afreen.jbasha@gmail.com

C. H. N. Santhosh Kumar
CSE Department, ANURAG Engineering College, Kodad, India
santhosh.ph10@gmail.com

Vedula Sitharamulu
Department of Computer Science and Engineering, GITAM School of Technology, Telangana, India
vsitaramu.1234@gmail.com (corresponding author)

B. N. V. Madhu Babu
CSE Department, Teegala Krishna Reddy Engineering College, Telangana, India
bnvmadhubabu2014@gmail.com

Hanumantha Rao Battu
Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Andhra Pradesh, India
hanuma9999@yahoo.com

K. Venkataramana
CSE (DS) Department, Malla Reddy Engineering College, Telangana, India
venkatramanakanakandi16@gmail.com

Received: 10 March 2025 | Revised: 9 April 2025 and 24 April 2025 | Accepted: 1 May 2025

 DOI: https://doi.org/10.48084/etasr.10898

ABSTRACT

Digital payment systems are increasingly used to complete everyday transactions; however, their digital nature exposes users to the risk of fraudulent activity, necessitating advanced detection techniques to ensure security. This study proposes a hybrid fraud detection model for digital wallet transactions by integrating Adversarial Autoencoders (AAE) and Gated Recurrent Units (GRU), combining AAE's ability to learns robust latent representations, and GRU's ability to capture temporal dependencies within transaction sequences. The proposed method outperforms existing approaches, achieving 99% accuracy, 99% recall, 98% F1-score, 99% precision, and 99.4% Area Under the Curve (AUC). By effectively reducing both false positives and false negatives, the model improves fraud detection and mitigates financial risks in digital transactions.

Keywords-deep learning; fraud detection; unsupervised learning

I. INTRODUCTION

Digital transformation has reshaped the way transactions are conducted, with electronic payment systems being widely adopted. While these systems offer enhanced speed and convenience, their increasing reliance on digital infrastructure has introduced critical vulnerabilities. As a result, they have become prime targets for cybercriminals employing tactics

such as malware attacks, transaction spoofing, and identity theft to illicitly access financial resources [1]. These growing threats highlight the urgent need for robust and adaptive fraud detection mechanisms to ensure the security and resilience of digital financial ecosystems.

Traditional fraud detection methods, often based on fixed rules or simple statistical models, struggle to adapt to the dynamic nature of modern fraud tactics [2] and lack the flexibility and intelligence to respond swiftly to novel threats, leaving systems exposed to exploitation [3]. On the other hand, Artificial Intelligence (AI) and Machine Learning (ML) have emerged as critical tools in combating fraud [4], with their ability to process vast datasets, uncover hidden patterns, and adapt to evolving behaviors. These properties make them particularly well-suited for fraud detection [5, 6].

Within ML, deep learning has shown exceptional performance, especially when working with complex, sequential data like transaction histories [7]. For example, Recurrent Neural Networks (RNNs) are particularly effective in capturing temporal patterns that may indicate fraudulent activity [8, 9]. Among RNN variants, the Gated Recurrent Unit (GRU) stands out for its balance between computational efficiency and modeling capability [10]. When combined with Adversarial Autoencoders (AAE), this pairing creates a powerful framework for fraud detection. AAEs are designed to capture the essential structure of normal data, making them well-suited for identifying deviations that signal fraud [11].

By integrating AAEs with GRUs, this study introduces an advanced fraud detection model that learns baseline transaction behavior and identifies anomalies more effectively [12]. The proposed system enhances digital payment security by significantly improving the detection of fraudulent transactions, emphasizing its real-world applicability [13].

II. LITERATURE REVIEW

To be effective in detecting fraudulent transactions, ML-based models must achieve a delicate balance between high accuracy, low false positive rates, and adaptability to evolving fraud strategies [14].

Supervised learning algorithms such as Decision Trees, Support Vector Machines (SVMs), and Random Forests (RF) are commonly used in fraud detection; however, their effectiveness is often limited by the inherent class imbalance present in labeled fraud datasets. For instance, while RF models typically achieve around 90% accuracy, their F1-score often drops below 70%, reflecting poor performance in detecting minority fraud cases [15]. In contrast, more advanced approaches such as the AAE-based model in [16] show more promise, achieving 94.5% accuracy and an 81.7% F1-score, with 85.2% precision and 78.6% recall. In another study focused on digital wallet fraud, the GRU outperformed both Long Short-Term Memory (LSTM) networks and Convolutional Neural Networks (CNNs), achieving 95.3% accuracy, 83.4% F1-score, 87.9% precision, and 79.8% recall, confirming its suitability for modeling sequential transaction data and its potential for secure financial applications [17]. Hybrid deep learning architectures have also shown strong performance, with the CNN-RNN model in [18], designed for real-time fraud detection, reporting 94.2% accuracy, 95.3% precision, 94.8% recall, and a 95.0% F1-score, along with an Area Under the Curve (AUC) of 96.1%.

III. PROPOSED METHODOLOGY

Figure 1 illustrates the architecture of the proposed hybrid AAE-GRU model. This approach follows a structured pipeline consisting of data preprocessing, feature extraction, model training, and fraud detection. By combining deep representation learning with sequential modeling, the framework significantly improves detection accuracy and adaptability in dynamic digital financial environments.

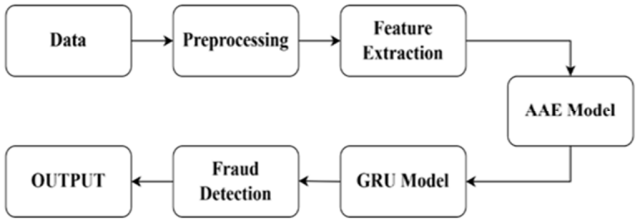

Fig. 1. Proposed architecture.

A. Data & Preprocessing

The "digital_wallet_transactions" dataset, obtained from a digital wallet system [19], comprises 5,035 samples, each labeled as either fraudulent or genuine. For the model development, a standard 80:20 train-test split was adopted, with 4,028 samples used for training and 1,007 for testing. Missing values were handled using K-Nearest Neighbors (KNN) imputation, while records with missing essential features were excluded. Categorical variables were one-hot encoded, and numerical features were normalized via Min-Max Scaling. To address skewness in variables such as transaction amounts, log transformation was applied. Given the class imbalance inherent in fraud detection datasets, several strategies are employed, including Synthetic Minority Over-sampling Technique (SMOTE), random under-sampling, and class weight adjustment.

B. Feature Extraction

Feature extraction is a critical component in fraud detection systems, transforming raw transactional data into meaningful representations that enhance model effectiveness. Key features considered in this study include Transaction ID, Timestamp, Sender and Receiver IDs, Transaction Amount, Transaction Type, Location, Device Information, and Fraud Labels. In addition to these raw attributes, derived features such as transaction frequency, average transaction amount per user, and preferred payment mode are computed to capture behavioral patterns and contextual anomalies. These enriched features play a pivotal role in distinguishing fraudulent activities from legitimate transactions by highlighting deviations from normal transactional behavior.

C. Adversarial Autoencoders (AAE) & Gated Recurrent Units (GRU) Model

In the hybrid model, the AAE component learns optimal latent representations of transaction data by enforcing a

structured distribution on the latent space. It consists of an encoder that maps input transaction data x to a latent representation z using a neural network:

$$z = f_\theta(x) = ReLU(W_1 x + b_1) \quad (1)$$

The decoder reconstructs the input data from the latent vector z through:

$$x = g_\varphi(z) = sigmoid(W_2 z + b_2) \quad (2)$$

The GRU, a powerful sequential model, captures temporal relationships in transaction sequences. At each time step, it updates its hidden state using the following mechanisms:

- Reset gate, which determines how much of the previous hidden state h_{t-1} should be forgotten:

$$r_t = \sigma(W_r[h_{t-1}, x_t]) \quad (3)$$

- Update gate, which controls the extent to which previous information is retained in the current state:

$$z_t = \sigma(W_z[h_{t-1}, x_t]) \quad (4)$$

- andidate hidden state, which computes a new hidden state proposal based on the reset gate:

$$\tilde{h}_t = tanh(W_h[r_t \odot h_{t-1}, x_t]) \quad (5)$$

- Final hidden state, which blends the previous state and the candidate state based on the update gate:

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \quad (6)$$

where σ denotes the sigmoid activation function, $tanh$ is the hyperbolic tangent function, and $\odot$ represents element-wise multiplication. The update gate z_t regulates the retention of historical information, while the reset gate r_t controls the forgetting of irrelevant past data.

IV. RESULTS AND DISCUSSION

The performance of the proposed AAE-GRU model was evaluated using standard classification metrics: accuracy, precision, recall, F1-score, and AUC. These metrics are defined as follows:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (7)$$

$$Precision = \frac{TP}{TF+FP} \quad (8)$$

$$Recall = \frac{TP}{TP+FN} \quad (9)$$

$$F1-score = 2\frac{Precision \cdot Recall}{Precision+Recall} \quad (10)$$

where TP denotes true positives, TN true negatives, FP false positives, and FN false negatives.

Figure 2 presents the AAE training loss curve over 100 epochs, showing rapid convergence. The initial loss (~0.10) declines sharply during the first 20 epochs and then stabilizes near zero, indicating effective learning and strong data reconstruction. However, a loss approaching zero can sometimes suggest overfitting or mode collapse, necessitating the use of validation loss for confirming generalization.

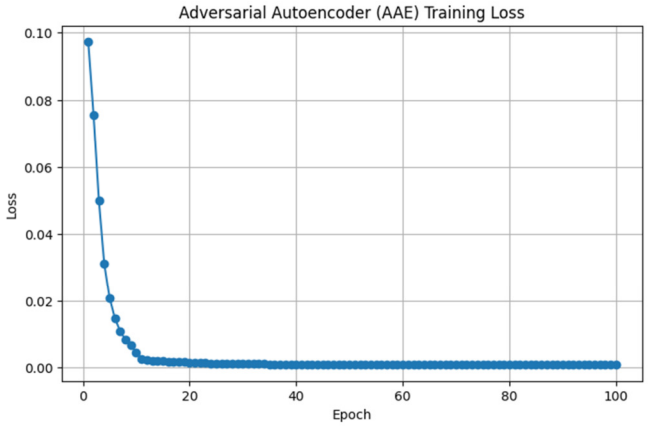

Fig. 2. AAE training loss.

The GRU model's training loss across 150 epochs, shown in Figure 3, steadily decreases from ~0.014 to ~0.004, reflecting successful sequential learning. Despite some oscillations typical of temporal data, the overall downward trend signifies robust training. Nonetheless, validation remains crucial to confirm model generalization and mitigate overfitting.

Figure 4 displays the confusion matrix for fraud detection, with the model correctly identifying 992 TP and 11 TN, alongside only 3 FP and 1 FN. This indicates high classification performance with minimal misclassifications. Importantly, the model's low false negative rate is critical for minimizing financial risk. Based on the confusion matrix, the error rate was calculated at only 0.397%.

Table I compares the performance of the AAE-GRU model against a CNN-RNN hybrid model on fraud detection tasks based on standard classification metrics. The AAE-GRU model outperforms its counterpart in every metric: 99% accuracy, precision, and recall, 98% F1-score, and 99.4% AUC, compared to approximately 95–96% for the CNN-RNN. These results highlight the efficiency and reliability of the proposed model in detecting fraudulent transactions.

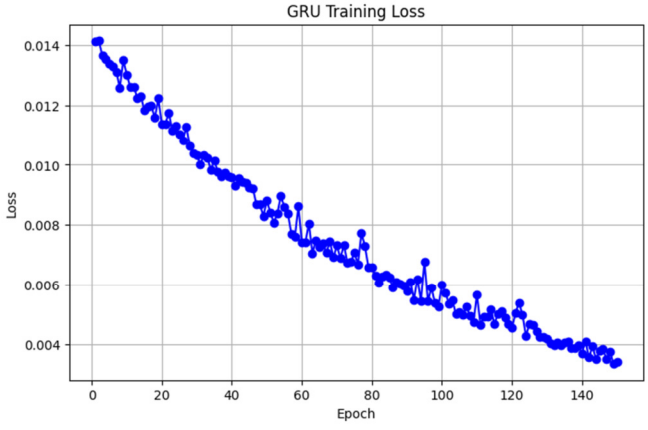

Fig. 3. GRU training loss.

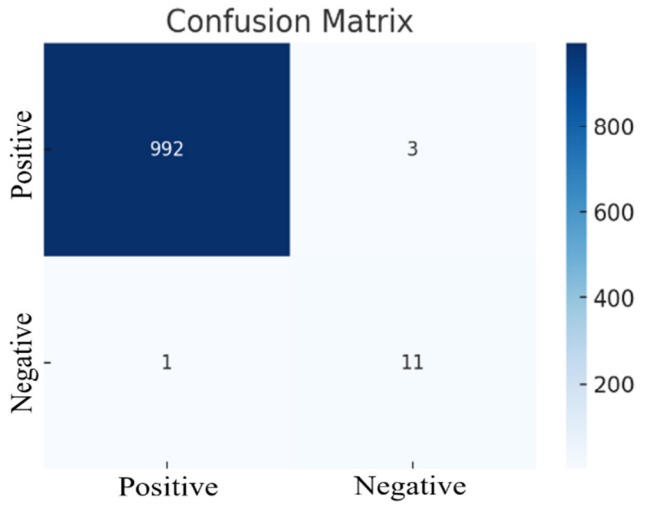

Fig. 4. Confusion matrix.

TABLE I. COMPARISON OF PERFORMANCE METRICS OF THE PROPOSED SYSTEM WITH A CNN-RNN MODEL

Performance Metrics	AAE-GRU	CNN-RNN
Accuracy	99%	95%
Recall	99%	95%
F1-score	98%	95%
Precision	99%	95%
AUC	99.4%	96.1%

Figure 5 illustrates the accuracy trends of the CNN-RNN and AAE-GRU hybrid models as the number of data samples increases from 1,000 to 5,035. The AAE-GRU model begins with an accuracy of 98% at 1,000 samples and reaches 99% at 5,035, while the CNN-RNN model improves from 92% to 95%. Throughout all sample sizes, the AAE-GRU consistently achieves higher accuracy and outperforms the CNN-RNN model.

Figure 6 compares the recall performance of the AAE-GRU and CNN-RNN hybrid models across varying sample sizes. The AAE-GRU model starts at 97.5% and increases to 99%, while the CNN-RNN model improves from 92% to 95%.

Figure 7 compares the F1-scores of the CNN-RNN Hybrid and AAE-GRU models across varying sample sizes. The AAE-GRU model starts at 97% and exceeds 98%, while the CNN-RNN model improves from 92% to 95%.

Figure 8 presents the precision of the CNN-RNN Hybrid and AAE-GRU models across different sample sizes. The AAE-GRU model begins at 97.5% and stabilizes around 99%, while the CNN-RNN model increases from 92% to 95%.

Figure 9 presents the AUC curve comparing the fraud detection performance of the CNN-RNN and AAE-GRU models. The AAE-GRU model achieves a higher AUC score of 99.4%, outperforming the CNN-RNN model, which records 96.1%.

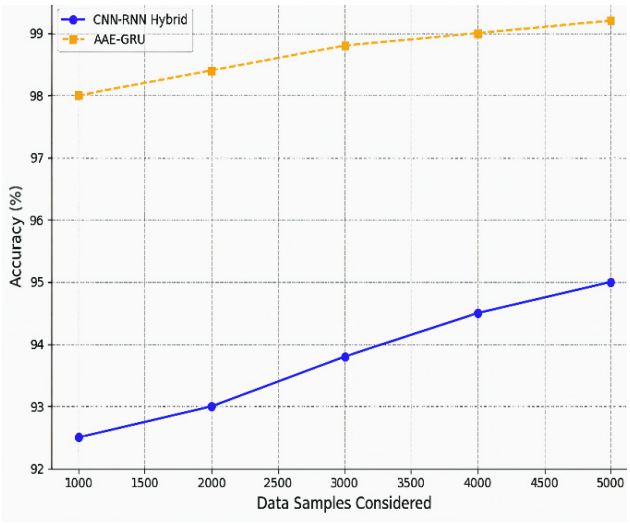

Fig. 5. Accuracy comparison of CNN-RNN and AAE-GRU.

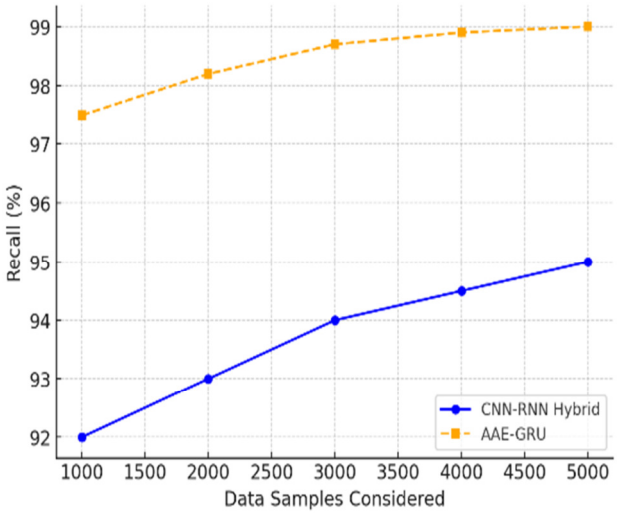

Fig. 6. Recall comparison of CNN-RNN and AAE-GRU.

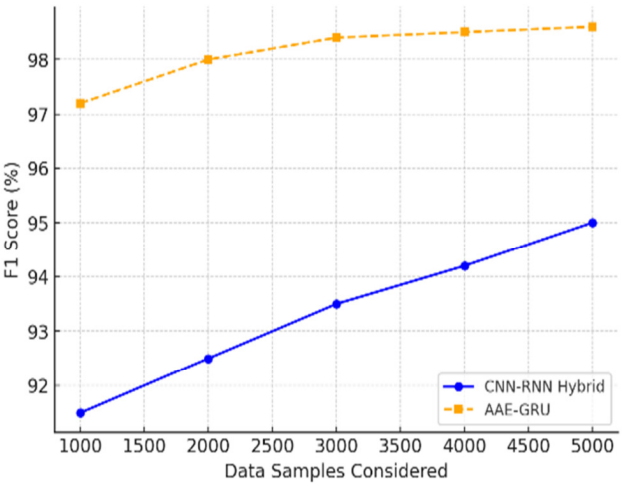

Fig. 7. F1-score comparison of CNN-RNN and AAE-GRU.

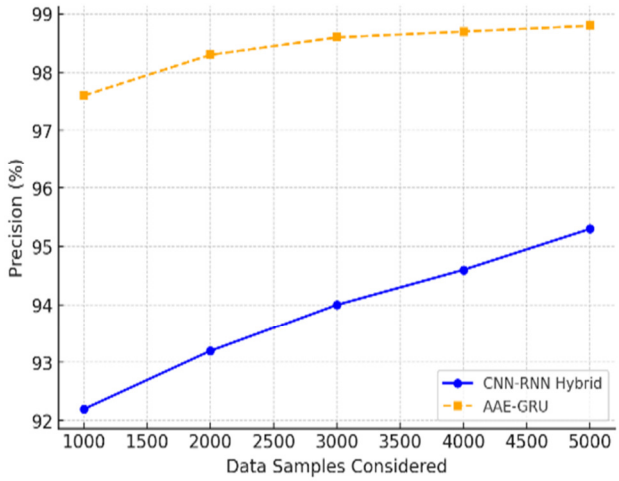

Fig. 8. Precision comparison of CNN-RNN and AAE-GRU.

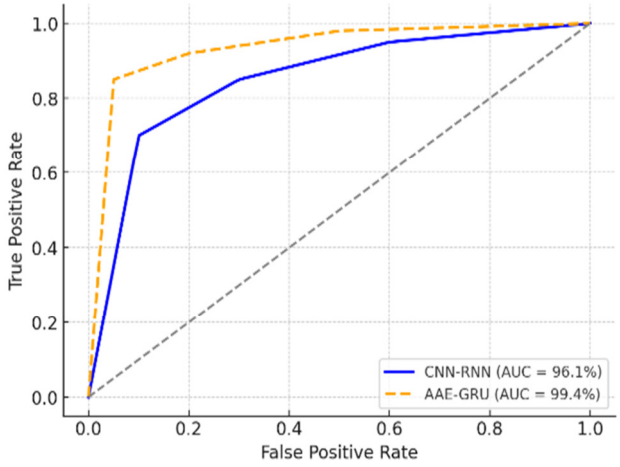

Fig. 9. AUC Curve comparison of CNN-RNN and AAE-GRU.

V. CONCLUSION

This study presents a novel fraud detection approach that integrates Adversarial Autoencoders (AAE) with Gated Recurrent Units (GRU). The proposed model effectively learns typical transaction patterns and accurately identifies anomalies, significantly reducing both false positives and false negatives. Its robustness and low false alarm rate make it a valuable tool for detecting fraud in digital wallet transactions.

Results reveal an overall improved performance of the proposed model across key evaluation metrics, achieving 99% accuracy, precision, and recall, 98% F1-score, and 99.4% Area Under the Curve (AUC), outperforming a Convolutional Neural Network-Recurrent Neural Network (CNN-RNN) hybrid counterpart model.

Future work will focus on enhancing the model's interpretability using techniques such as Shapley Additive Explanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME) to foster trust and adoption. Optimizing the model for real-time deployment will support integration into live transaction monitoring systems, enabling immediate fraud prevention. Additionally, exploring cross-domain adaptability through transfer learning could extend the model's applicability to other financial fraud domains, such as credit card fraud, thereby increasing its versatility and impact.

REFERENCES

[1] P. Majhi, "Fraud Detection in Financial Transactons," *Indian Scientific Journal of Research in Engineering and Management*, vol. 09, no. 01, pp. 1–9, Jan. 2025, https://doi.org/10.55041/ijsrem41105.

[2] P. Jeyachandran, A. S. V. V. Akisetty, P. Subramani, O. Goel, D. S. P. Singh, and E. A. Shrivastav, "Leveraging Machine Learning for Real-Time Fraud Detection in Digital Payments," *Integrated Journal for Research in Arts and Humanities*, vol. 4, no. 6, pp. 70–94, Nov. 2024, https://doi.org/10.55544/ijrah.4.6.10.

[3] F. G. Abdiwi, "Detection of Digital Currency Fraud through a Distributed Database Approach and Machine Learning Model," *TEM Journal*, pp. 3025–3039, Dec. 2024, https://doi.org/10.18421/TEM134-37.

[4] M. N. M. Sunny, K. M. S. Hossain, M. M. Amin, S. N. Sadmani, and M. A. Siddique, "Numerical analysis of multivariate data for fraud detection," *Nanotechnology Perceptions*, pp. 325–335, Dec. 2024, https://doi.org/10.62441/nano-ntp.vi.3486.

[5] K. Raghuveer, T. N. Gongada, D. Nimma, M. Aarif, K. Samudrala, and B. K. Bala, "Enhancing Fraud Detection in Online E-Commerce Transactions through Deep Learning Auto encoder Model," in *2024 International Conference on Intelligent Systems and Advanced Applications (ICISAA)*, Pune, India, Oct. 2024, pp. 1–6, https://doi.org/10.1109/ICISAA62385.2024.10828858.

[6] A. A. Alhashmi, A. M. Alashjaee, A. A. Darem, A. F. Alanazi, and R. Effghi, "An Ensemble-based Fraud Detection Model for Financial Transaction Cyber Threat Classification and Countermeasures," *Engineering, Technology & Applied Science Research*, vol. 13, no. 6, pp. 12433–12439, Dec. 2023, https://doi.org/10.48084/etasr.6401.

[7] V. Anitha, "A Survey on Online Payment Fraud Detection Techniques using Machine Learning Algorithms," *International Journal for Research in Applied Science and Engineering Technology*, vol. 13, no. 1, pp. 1003–1010, Jan. 2025, https://doi.org/10.22214/ijraset.2025.66490.

[8] Y. R. Maramreddy and K. Muppavaram, "Detecting and Mitigating Data Poisoning Attacks in Machine Learning: A Weighted Average Approach," *Engineering, Technology & Applied Science Research*, vol. 14, no. 4, pp. 15505–15509, Aug. 2024, https://doi.org/10.48084/etasr.7591.

[9] A. Sohel, M. A. Alam, M. Waliullah, A. Siddiki, and M. M. Uddin, "Fraud Detection in Financial Transactions Through Data Science for Real-Time Monitoring and Prevention," *Academic Journal on Science, Technology, Engineering & Mathematics Education*, vol. 1, no. 01, pp. 91–107, Oct. 2024, https://doi.org/10.69593/ajieet.v1i01.132.

[10] H. P. Tarigan, "Development Of Machine Learning Algorithms For Fraud Detection In Digital Transactions," *Jurnal Komputer Indonesia*, vol. 3, no. 1, Jul. 2024, https://doi.org/10.37676/jki.v3i1.571.

[11] G. Rajeshwari, S. Mownika, G. Anupriya, and R. Kishore, "Fraud Detection in E-Commerce Transactions Using Machine Learning Techniques and Quantum Networks:," in *Quantum Networks and Their Applications in AI*, C. Ananth, O. Ibrahim Khalaf, and J. Anand, Eds. IGI Global, 2024, pp. 146–162.

[12] V. Kant, "Optimizing Logistic Regression for Flawless Fraud Detection in Digital Payments," in *2024 Second International Conference on Intelligent Cyber Physical Systems and Internet of Things (ICoICI)*, Coimbatore, India, Aug. 2024, pp. 97–100, https://doi.org/10.1109/ICoICI62503.2024.10696469.

[13] S. Kumar, "Enhanced Fraud Detection in Financial Transactions Using Hyperparameter-Tuned Random Forests," in *2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, Kamand, India, Jun. 2024, pp. 1–7, https://doi.org/10.1109/ICCCNT61001.2024.10725958.

[14] J. N. Benedict, V. S. Kiran, G. R. Nivetha, and S. Senthil Pandi, "A Novel Multi-Factor Graphical Authentication Scheme for Enhanced Digital Security," in *2024 International Conference on Computational*

Intelligence for Green and Sustainable Technologies (ICCIGST), Vijayawada, India, Jul. 2024, pp. 1–6, https://doi.org/10.1109/iccigst60741.2024.10717545.

[15] M. Farouk *et al.*, "Fraud_Detection_ML: Machine Learning Based on Online Payment Fraud Detection," *Journal of Computing and Communication*, vol. 3, no. 1, pp. 116–131, Feb. 2024, https://doi.org/10.21608/jocc.2024.339929.

[16] N. N. Jose, A. K. Arigela, G. Vivekanandan, R. Sethuraman, S. B. T. Naganathan, and N. Venu, "Optimizing Payment Transaction Security: Utilizing Gradient Boosting Machines for Fraud Detection," in *2024 10th International Conference on Communication and Signal Processing (ICCSP)*, Melmaruvathur, India, Apr. 2024, pp. 720–725, https://doi.org/10.1109/ICCSP60870.2024.10543774.

[17] J. Lu, "Improving Fraud Detection in Mobile Payments with Machine Learning Ensembles," in *2024 4th International Conference on Electronic Information Engineering and Computer Science (EIECS)*, Yanji, China, Sep. 2024, pp. 851–854, https://doi.org/10.1109/EIECS63941.2024.10800126.

[18] S. Lenka and R. Tiwari, "Real-Time Fraud Prevention in Digital Wallet Transactions Using CNN-RNN Hybrid Networks," *Cuestiones De Fisioterapia*, vol. 54, no. 2, pp. 533–542, 2025, https://doi.org/10.48047/cu.

[19] *Digital Wallet Transactions.* (2025), H. Rai. [Online]. Available: https://www.kaggle.com/datasets/harunrai/digital-wallet-transactions.