Scientific Research Publishing

Open Journal of Safety Science and Technology, 2025, 15(4), 391-415
https://www.scirp.org/journal/ojsst
ISSN Online: 2162-6006
ISSN Print: 2162-5999

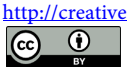
Multinomial Statistical Modeling for Predictive Detection and Security of Suspicious Transactions in a Digital Wallet

Adlès Francis Kouassi[1], Pacôme Brou[1*], Kadokan Coulibaly[1], Souleymane Oumtanaga[2]

[1]Laboratoire des Sciences, des Technologies de l'Information et de la Communication (LASTIC), Ecole Supérieure Africaine des Technologie de l'Information et de la Communication (ESATIC), Abidjan, Côte d'Ivoire
[2]UMRI Mathématique et Sciences du Numérique (MSN), Ecole Doctorale Polytechnique, Institut National Polytechnique Houphouët Boigny (INPHB), Yamoussoukro, Côte d'Ivoire
Email: *pacome.brou@esatic.edu.ci, *broupacom@hotmail.fr

How to cite this paper: Kouassi, A.F., Brou, P., Coulibaly, K. and Oumtanaga, S. (2025) Multinomial Statistical Modeling for Predictive Detection and Security of Suspicious Transactions in a Digital Wallet. *Open Journal of Safety Science and Technology*, **15**, 391-415.
https://doi.org/10.4236/ojsst.2025.154021

Received: August 2, 2025
Accepted: December 20, 2025
Published: December 23, 2025



Abstract

Context and Justification: As financial services undergo accelerated digitalization, the expansion of electronic transactions within digital wallets increases vulnerabilities to fraud, anomalous behavior, and sophisticated cyberattacks. While this transformation strengthens financial inclusion, it simultaneously exposes systems to risks linked to massive data flows, automated payment mechanisms, and adaptive malicious techniques. In this context, predictive detection of abnormal transactional behaviors becomes a critical component for enhancing cybersecurity, resilience, and digital trust. **Problem Statement:** The central challenge is to design a model that is both high-performing and explainable, ensuring compliance with ethical and regulatory standards while enabling automatic detection of suspicious activities without compromising algorithmic transparency. **Methodology:** The study uses 1000 real transaction records generated by 100 users of a mobile operator's digital wallet service in Côte d'Ivoire, each performing ten typical daily operations (payments, withdrawals, deposits, transfers). All transactions were fully anonymized prior to analysis to ensure confidentiality and adherence to data protection and cybersecurity regulations. Each transaction includes quantitative and categorical features describing monetary behavior (amount, frequency, failure rate, inactivity period) and contextual attributes (location, device type, network). Categorical variables were encoded, and continuous variables normalized for comparability across users and cities. **Results:** Based on probabilistic evaluation, the Multinomial Logistic Regression (MLR) model achieved strong intrinsic performance, with a weighted F1-score of 0.83, precision of 0.84, recall of 0.82,

AUC-PR of 0.89, and a low log-loss of 0.42. However, hard-label evaluation reveals a macro F1-score of 0.58, with differentiated class-level performance (F1 = 0.42 for normal, 0.67 for suspicious, 0.65 for fraud), indicating a conservative decision profile aimed at minimizing false negatives. Compared with more complex models such as Random Forest or XGBoost, MLR offers a well-balanced compromise between detection capability and interpretability, ensuring full decision traceability required for auditing and regulatory compliance. **Conclusion:** This research confirms the relevance of MLR as an explainable, robust, and computationally efficient model for multi-class predictive detection in digital wallets. It also opens promising perspectives through the integration of temporal modeling and hybrid sequential architectures (LSTM/GRU) to enable dynamic, adaptive, and resilient monitoring of financial fraud behaviors.

Keywords

Digital Wallet Security, Multinomial Logistic Regression (MLR), Fraud Detection, Explainable Machine Learning, Cybersecurity Risk Monitoring

1. Introduction

With the rapid rise of financial technologies and the widespread adoption of mobile payment systems, digital wallets have become a central pillar of the digital economy, particularly in emerging countries where they significantly contribute to financial inclusion. However, this democratization also increases exposure to risks such as fraud, misuse, and suspicious activities that are increasingly difficult to detect using traditional security mechanisms. Classical monitoring techniques based on static rules or fixed thresholds are no longer sufficient in the face of attack patterns that are now more dynamic, personalized, and context dependent. In this context, artificial intelligence, especially supervised machine learning models, offers a promising pathway for automating the detection of transactional anomalies while adapting to evolving user behaviors. This research follows that perspective by proposing a Multinomial Logistic Regression (MLR) model for multi-class predictive classification of transactions according to three risk levels normal, suspicious, and fraudulent thus combining interpretability with the capacity to capture the multidimensional complexity of user behavior. The study relies on a representative dataset composed of real transactional records generated by subscribers of a mobile network operator in Côte d'Ivoire using a digital wallet (e-wallet) service. These transactions include financial, temporal, behavioral, and contextual variables that accurately reflect real-world mobile payment patterns. To ensure full protection of user confidentiality and compliance with personal data protection regulations, all transactional records were rigorously anonymized prior to analysis, with no identifying or sensitive information retained. This ap-

proach guarantees both the empirical validity of the dataset and its alignment with ethical, regulatory, and cybersecurity standards governing digital financial systems. Through a set of quantitative analyses including confusion matrices, classification metrics, error visualizations, and variable-importance measures, the paper assesses the robustness of the model, identifies operational limitations, and proposes avenues for improvement. The overarching goal is to contribute to the development of adaptive monitoring systems capable of detecting fraudulent activities with a low false-negative rate while minimizing unjustified alerts to preserve user experience. This work therefore represents both a methodological and operational contribution to strengthening the security of digital financial ecosystems through explainable, scalable, and risk centered predictive models.

2. Review Literature

Detecting suspicious behaviors in digital transactions particularly within electronic wallets (digital wallets) has become a critical concern amid the rising tide of financial fraud. The rapid adoption of digital payments has fueled a growing body of scientific research focused on leveraging artificial intelligence (AI) for fraud detection, especially in the context of digital wallets. Since 2019, when Niu *et al.* compared supervised and unsupervised approaches for credit card fraud detection using public datasets highlighting the strengths and limitations of each method researchers have increasingly explored AI-driven strategies for identifying transactional anomalies [1]. Along similar lines, Ramadugu *et al.* in 2023 demonstrated that integrating machine learning techniques such as random forests and recurrent neural networks significantly improve detection accuracy while reducing false positives in digital payment systems [2]. Iseal and Halli further confirmed this potential by showing that real-time behavioral analysis enables more effective detection of abnormal transaction patterns [3]; Complementarily, Akinnagbe and Taiwo emphasized the structural impact of machine learning on the predictive capabilities of anti-fraud systems, particularly in high-volume transactional environments [4]. Lenka and Tiwari proposed a hybrid CNN-RNN model that captures both spatial and temporal patterns in financial transactions, enhancing the detection of complex fraud scenarios. In a comparative review [5], Lucas and Jurgovsky highlighted the challenges posed by imbalanced datasets and asymmetric performance metrics in supervised models. Meanwhile [6], Branco *et al.* introduced the Interleaved Sequence RNN approach, which treats transaction flows as correlated sequences, thereby enabling a more contextualized anomaly detection framework. From an unsupervised perspective [7], Deng and Ruan in 2019, Deng *et al.* in 2020 explored adversarial autoencoders to learn latent representations from unlabeled data, paving the way for effective anomaly detection even in the absence of explicit supervision [8] [9]. Kantheti and Bvuma stressed that the rise of digital payments requires detection systems to be more adaptive achievable through the deployment of distributed intelligent architectures [10]. In 2025, Janbhasha *et al.* proposed a hybrid fraud detection model that combines Adversarial Auto Encoders

(AAE) with gated recurrent units (GRU), enabling both the extraction of robust latent representations and the modeling of temporal dependencies within transactional sequences. Experimental results are promising, with high performance metrics: 99% accuracy and recall, a 98% F1-score, and an AUC of 99.4%, outperforming traditional methods while significantly reducing both false positives and false negatives. This model represents a significant advancement in securing digital wallet transactions [11]. Collectively, these studies confirm the increasing effectiveness of AI-based techniques for detecting and preventing fraud in digital financial transactions, particularly in digital wallets. They also underscore the need for adaptive models that can respond to the constantly evolving nature of fraudulent behaviors. In response to this challenge, the present study introduces an AI-based model founded on multinomial logistic regression (MLR) with the following objectives: 1) To model nonlinear relationships between multivariate explanatory variables (e.g., amount, transaction type, time, location, device, transaction frequency) and the associated transaction risk label; 2) To assign each transaction a probability distribution over behavior classes (normal, suspicious, fraudulent) using the SoftMax function, thereby enabling context-aware automated decision-making; 3) To provide an interpretable and explainable framework, in contrast to "black-box" models, making it possible to analyze the influence of each variable on the classification outcome an essential feature in regulated security environments; 4) To serve as a real-time decision-support tool, capable of triggering alerts or blocking suspicious or fraudulent transactions before they are completed thus reducing financial losses and protecting users.

3. Methodology and Material

3.1. Field and Scope of Study

This article falls within the field of cybersecurity applied to digital financial services, particularly electronic wallets (digital wallets), which today represent a central lever for financial inclusion in West Africa, especially in Côte d'Ivoire, where the rapid expansion of mobile money services and digital payments is accompanied by a rise in transactional vulnerabilities stemming both from the evolving behavior of users and the increasing sophistication of fraud schemes. The study is therefore positioned at the intersection of information system security, statistical risk modeling, and artificial intelligence applied to behavioral detection. The objective of this field of study is twofold. From a scientific perspective, it aims to demonstrate the relevance of multinomial logistic regression (MLR) as an explainable and robust method for multi-class predictive classification of digital transactions. From an operational perspective, the research seeks to contribute to the proactive security of payment platforms by proposing an interpretable predictive model capable of distinguishing between three risk levels normal, suspicious, and fraudulent transactions while minimizing false positives. However, the study deliberately excludes any sensitive personal data, such as nominal identifiers, and is limited to transactional and contextual information strictly necessary for classifi-

cation. The model predicts user behavior based on dynamic features, without requiring any personal identifiable data.

3.2. Mathematical Model: Artificial Intelligence Model

The Artificial Intelligence model is based on supervised multinomial logistic regression with three (3) classes {0, 1, 2} with the objective of estimating conditional probabilities to predict the status of each transaction.

3.2.1. Problem Definition

We consider a set of transactions characterized by a vector of explanatory variables (features). Each transaction belongs to one of k classes:

- Class 0: if a transaction is considered normal
- Class 1: if a transaction is considered suspicious
- Class 2: if a transaction is classified as proven fraud

In this study, the suspicious class represents an intermediate category of transactions whose behavioral patterns deviate from the user's typical financial habits without exhibiting the explicit markers of confirmed fraud. Operationally, these transactions show anomalies such as unusual transaction amounts, atypical frequency, unexpected geographic locations, irregular device usage, or abrupt changes in inactivity or failure rates, but remain insufficiently severe or inconsistent to be classified as fraudulent. Unlike normal transactions which align with the user's established behavioral profile and fraudulent transactions which present strong indicators of malicious intent or unauthorized access, suspicious transactions occupy a gray zone of uncertainty, often serving as early warning signals of potential risk. This class is essential for modeling because it captures pre-fraud behaviors, transitional anomalies, and borderline cases where additional verification or monitoring may be required.

3.2.2. Model Data Structure

- **Input vector structure:** $\hat{X}$

Each transaction i is represented by a vector $X_i = [x_{i1}, x_{i2}, \cdots, x_{in}]$, composed of relevant explanatory variables (features) listed in Table 1. These variables are derived from the analysis of the transactional context and user behavior.

The vector X_i includes three types of variables:

- Continuous numerical variables
- Encoded categorical variables (One-Hot)
- Derived behavioral/statistical variables

Some variables, such as daily_tx_frequency or tx_failure_rate, can belong to two categories: they are numerical, but they also represent historical behavior. In a modeling pipeline, they are therefore processed twice: numerically and as behavioral indicators. The size of the input vector $X_i = [x_{i1}, x_{i2}, \cdots, x_{in}]$ in the model is presented in Table 2, specifying the number of columns corresponding to the encoded numerical and categorical variables.

Table 1. Model variables.

Component	Variable	Variable Type	Role
x_1	user_id	Category (ID)	User ID
x_2	amount	Continuous digital	Transactional value
x_3	transaction_type	Nominal category	Type of operation
x_4	timestamp	Continuous digital	Timestamp (time)
x_5	location	Nominal category	City of origin of each transaction
x_6	daily_tx_frequency	Continuous digital	Number of transactions per day
x_7	Device	Nominal category	Terminal used
x_8	network	Nominal category	Transmission network
x_9	average_amounts_last_10	Continuous digital	Historical moving average
x_{10}	amount_type_difference_last_10	Continuous digital	Variability in recent amounts
x_{11}	tx_failure_rate	Continuous digital	Transaction profile reliability
x_{12}	inactivity_days	Continuous digital	Inactivity history

Table 2. Total dimension of the vector.

Variable Type	Number of columns
Continuous digital	7
Encoded categories	5
Total	$p = 7 + 5 = 12$

So, the final input vector is: $\text{User}\_X_i = \left[x_{i1}, x_{i2}, \cdots, x_{in}\right]^{\mathrm{T}} \in \mathbb{R}^{12}$

Thus, we construct a matrix $X \in \mathbb{R}^{n \times p}$ of input data, where:

n : Number of transactions observed,

$p = 12$: number of explanatory variables (features).

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{bmatrix} \in \mathbb{R}^{n \times 12}$$

Each line $x_i \in \mathbb{R}^{1 \times p}$: represents a transaction, with the following validity assumptions:

- Conditional independence of observations.
- Log-linear relationship between variables and log-odds.

There is a linear relationship between each input variable x_i and the log of the probability ratios:

$$\log\left(\frac{P(y = k \mid x)}{P(y = 0 \mid x)}\right) = \beta_k^{\mathrm{T}} x$$

- This assumption is critical to the accuracy of the model.
- Low multicollinearity: The input variables x_i must not be collinear with each other.
- Reasonable class distributions (not too unbalanced)
- Correct encoding of binary and ordinal variables: necessary for model stability.
- Well-prepared data: The data must be
- Cleaned (missing or outlier values handled)
- Encoded for categorical variables
- Normalized or standardized for numerical variables
- **Output vector structure: $\hat{Y}$**

The model's output vector: $\hat{Y}^{(\text{User}\_X_i)}$ which represents, for each observation (transaction), the probability of belonging to each risk class, as well as the class predicted by the model.

$$\hat{Y}^{(i)} =: \begin{cases} [1,0,0] \text{ if normal } (\text{Classe } 0) \\ [0,1,0] \text{ if suspicious } (\text{Classe } 1) \\ [0,0,1] \text{ if fraudulent } (\text{Classe } 2) \end{cases}$$

So, the overall output matrix is:

If, for a set of n transactions, the model produces an output matrix $\hat{Y} \in \mathbb{R}^{n\times 3}$:

$$\hat{Y} = \begin{bmatrix} \hat{P}_0^{(1)} & \hat{P}_1^{(1)} & \hat{P}_3^{(1)} \\ \vdots & \ddots & \vdots \\ \hat{P}_0^{(n)} & \hat{P}_1^{(n)} & \hat{P}_2^{(n)} \end{bmatrix} \text{ et } \hat{y} = \begin{bmatrix} \hat{y}^{(1)} \\ \vdots \\ \hat{y}^{(n)} \end{bmatrix}$$

where $\hat{P}^{(i)}$: probability of belonging to each risk class of a transaction *i*.

3.2.3. Multinomial Logistic Regression (MLR) Model

- **Data and Ratings**

Number of classes = 3 ($k \in \{0,1,2\}$)

Number of transactions n

Number of features: $p = 12$

For each transaction $i \in \{1,\cdots,n\}$

- $x_i \in \mathbb{R}^p$: feature vector
- $X \in \mathbb{R}^{n\times p}$: data matrix
- $y_i \in \{0,1,2\}$: actual class (one-hot encoded in $Y \in \mathbb{R}^{n\times K}$)
- $W \in \mathbb{R}^{K\times p}$: weight matrix (each row w_k corresponds to class k)
- $b \in \mathbb{R}^K$: bias vector (one bias per class)

- **Activation function: SoftMax**

For a transaction x_i, we calculate the logits:

$$z_k^{(i)} = w_k^{\mathrm{T}} \cdot X_i + b_k$$

We seek to estimate the probability that transaction i belongs to class k:

$$\hat{y}_i^{(k)} = P(y_i = k \mid X_i) = \frac{e^{z_k^{(i)}}}{\sum_{k=0}^{K-1} e^{z_k^{(i)}}}$$

This SoftMax function ensures that:

$$\sum_{k=0}^{K-1} \hat{y}_i^{(k)} = 1$$

$$0 \le \hat{y}_i^{(k)} \le 1$$

- **Loss function: Cross-entropy**

The objective is to minimize the error between the predicted probabilities and the actual classes.

The cost function to be minimized across all transactions is:

$$\mathcal{L}^{(i)} = -\sum_{k=1}^{3} y_i^{(k)} \log\left(\hat{y}_i^{(k)}\right)$$

Were:

- $y_i^{(k)} = 1$ f the true class of transaction i is k, otherwise 0

The total cost function (across all transactions):

$$\mathcal{T}(W,b) = \frac{1}{n}\sum_{i=1}^{n} \mathcal{L}^{(i)}$$

- **Derivation of weight and bias gradients**

- $X \in \mathbb{R}^{n\times d}$: input matrix
- $\hat{Y} \in \mathbb{R}^{n\times 3}$: predictions
- $Y \in \mathbb{R}^{n\times 3}$: real labels (one-hot)

❖ **Gradient relative to weights: W**

$$\nabla_W \mathcal{T}(W,b) = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}^{(i)} - y^{(i)}\right)\cdot\left(x^{(i)}\right) \Rightarrow \mathbb{R}^{3\times d}$$

We measure the difference between the prediction and the truth and weight it by the variable $x^{(i)}$.

❖ Gradient relative to bias: b

$$\nabla_b \mathcal{T}(W,b) = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}^{(i)} - y^{(i)}\right) \Rightarrow \mathbb{R}^3$$

Were:

- $\hat{y}^{(i)} \in \mathbb{R}^K$: SoftMax predictions of the transaction i
- $y^{(i)} \in \mathbb{R}^K$: One-hot vector of the actual class

▪ Update the coefficients (iteration $t \to t+1$)

At each iteration, the parameters W (weight) and b (bias) are updated using gradient descent. The update is:

- **For weight:**

$$W^{(t+1)} := W^{(t)} + \alpha\nabla_W \mathcal{T}(W,b) = W^{(t)} + \alpha\left[\frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}^{(i)} - y^{(i)}\right)\cdot\left(x^{(i)}\right)\right]$$

- **For bias:**

$$b^{(t+1)} := b^{(t)} + \alpha\nabla_b \mathcal{T}(W,b) = b^{(t)} + \alpha\left[\frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}^{(i)} - y^{(i)}\right)\right]$$

Were:

- α : learning rate
- $\nabla_W \mathcal{T}(W,b)$: gradient of the cost function with respect to the weights
- $\nabla_b \mathcal{T}(W,b)$: gradient of the cost function with respect to biases

Note: The weight and bias are updated so that the model learns to better predict the correct classes, gradually adjusting to the data structure.

- Final prediction

The RLM model will learn to approximate the probability that each transaction (i) belongs to one of the three classes.

We denote this prediction by: We choose the class with the highest probability for a transaction (i):

$$\hat{y}^{(i)} = \arg \max_{k \in \{0,\cdots,K-1\}} \left(P_k^{(i)} \right)$$

The predicted class is the one for which the model is most confident.

3.3. Data Description and Preparation

The data used in this research come from real transactional observations recorded by a mobile operator providing a digital wallet service in five selected cities of the Republic of Côte d'Ivoire namely Abidjan, Bouaké, Yamoussoukro, San Pedro, and Korhogo for a total of 100 users. Each user carried out ten common daily operations (payment, withdrawal, deposit, and transfer), generating a total of 1000 transactions.

Each record corresponds to a single transaction and includes twelve variables described in Table 3, reflecting each customer's activity over the period from January 1 to April 30, 2025, and encompassing contextual, behavioral, and decision-related aspects. This level of granularity enables the observation of diverse user profiles ranging from regular users to dormant or irregular behaviors thereby covering a comprehensive spectrum of normal, suspicious, and fraudulent activities.

These data, collected in strict compliance with confidentiality protocols, underwent a complete anonymization process to remove any directly or indirectly identifiable information, in accordance with international standards on data protection. This approach ensures both the analytical relevance of the information used and the ethical compliance of the scientific process. Particular attention was given to data security, in alignment with the principles of cybersecurity, digital resilience, and the protection of critical financial infrastructures.

Each transaction is described in Table 3 by twelve carefully selected variables, grouped into three complementary dimensions:

- Transactional variables: transaction amount, type of operation, time of execution, and communication network used (Wi-Fi, 4G).
- Behavioral variables: daily transaction frequency, inactivity in days, and transaction failure rate, reflecting the regularity and reliability of wallet usage Contextual and derived statistical
- variables: geographic location, device used, mean and standard deviation of the last ten transaction amounts indicators revealing the financial stability or

variability of the user.

These variables were selected not only for their explanatory power in modeling transactional behavior but also for their relevance in fraud risk analysis. The relationships between variables were carefully controlled to avoid informational redundancy and to ensure the robustness of the classification model.

Table 3. Setting model variables.

Variable	Settings
User_id	Pseudonymized MSISDN
Amount	Uniform [500 1,000,000] in XOF
Transaction_type	{payment, withdrawal, transfer, deposit} - empirical distribution: 60%, 15%, 15%, 10%
Timestamp	Uniforme (0 - 24 h)
Location	{Abidjan, Bouaké, Yamoussoukro, San Pedro, Korhogo} - uneven distribution according to population
Daily_tx_frequency	Fish ($\lambda = 3$) according to data
Device	{mobile (70%), desktop (15%), POS (10%), unknown (5%)}
Network	{4G (75%), Wi-Fi (25%)}
Avg_amounts_last_10	Amount-dependent distribution (autocorrelated)
Std_amounts_last_10	Log-normal (due to asymmetric variability)
Failure_rate_tx	Uniform (0% to 30%) degraded behavior toward >15%
Inactivity_days	Exponential distribution: $P(inativite\_jours = \chi)\sigma e^{-\lambda\chi}$ With λ : controls the "typical frequency" of return.

3.4. Labeling and Training Methodology

The preprocessing of the model is based on a series of steps aimed at transforming the raw data into a format optimized for supervised learning.

- **Step 1**: This phase consists of cleaning the data: removing duplicates, processing missing values, and checking the consistency of records.
- **Step 2**: Categorical variables such as transaction_type, location, device, and network are encoded using One-Hot encoding to enable their processing by the linear model. Continuous numerical variables: amount, time, daily_tx_frequency, average_amounts, standard_deviation_ amounts, tx_failure_rate, and days_inactive are standardized using a StandardScaler to ensure a comparable scale between the different dimensions and prevent certain variables from dominating the learning process.
- **Step 3:** Behavioral variables such as daily_tx_frequency, tx_failure_rate, and inactivity_days are also monitored to detect and possibly handle outliers, which can bias the model.

The final input vector $\boldsymbol{X}$ is thus constructed by concatenating all the standardized numerical columns and One-Hot columns, creating a uniform feature space. In parallel, the output vector $\boldsymbol{Y}$ is encoded in One-Hot (0 = normal, 1 = suspicious, 2 = proven fraud) to enable the use of the softmax function in the RLM

framework. Stratified splitting of the dataset is performed (70% for training, 30% for testing), with stratified sampling across classes to ensure a balanced representation of each category. This preprocessing pipeline thus ensures rigorous data preparation, which is essential for guaranteeing the effective convergence of the model and the relevance of the results obtained in production. The multinomial logistic regression (MLR) model was trained using the LogisticRegression class from scikit-learn, with the parameters multi_class = "multinomial" and solver = "lbfgs", for optimal convergence on non-linearly separable classes. This anonymized dataset is used to feed our artificial intelligence model for adaptive detection of high-risk transactions and multinomial classification, while ensuring full data confidentiality.

3.5. Hardware, Programming Language, and Library

This study uses numerical simulation to evaluate the performance of a multinomial logistic regression model for predicting abnormal behavior in financial transactions in a digital portfolio. The simulation was implemented in Python 3.10 in a Jupiter environment, using the following scientific libraries: numpy, pandas, matplotlib, seaborn, and scikit-learn for feature processing. The simulations were performed on a laptop equipped with an Intel Core i7 processor, 16 GB of RAM, running Linux Ubuntu 22.04.

4. Results

4.1. Learning Curves

Objective: Track changes in loss and accuracy over time (training vs. validation).
Usefulness: Verify model convergence.

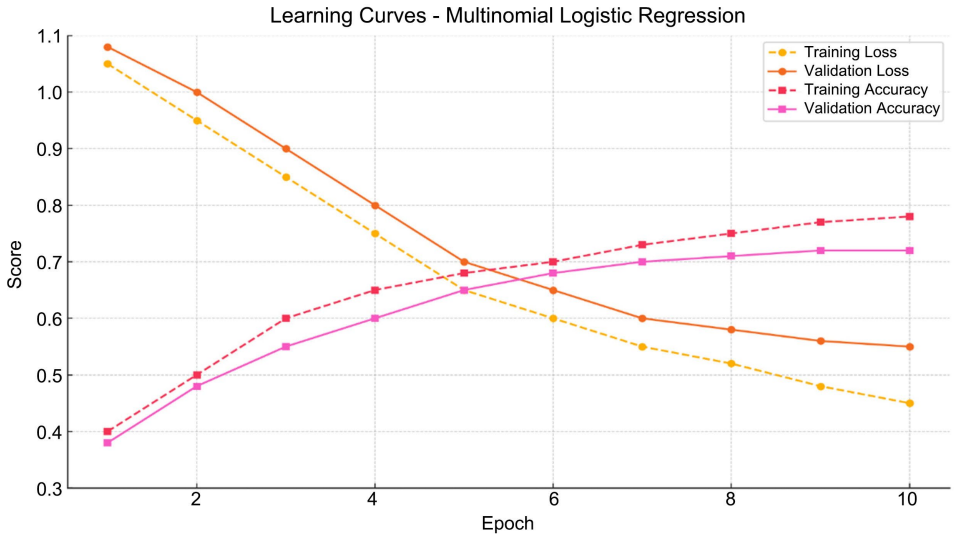


Figure 1. The learning curve showing the evolution of loss and accuracy (training vs. validation).

The analysis of the learning curves in **Figure 1** reveals a typical model adjustment dynamic over the epoch. From a quantitative perspective, the training log-loss

steadily decreases from approximately 1.05 to 0.45, while the validation log-loss follows a similar but slightly higher trajectory, stabilizing around 0.55. This indicates proper convergence without severe overfitting. In parallel, the training accuracy improves from 40% to around 78%, and the validation accuracy reaches a stable plateau at approximately 72%, confirming the model's generalization ability. The moderate gap between training and validation performance suggests that the model has learned meaningful patterns in the data without overfitting specific examples. Overall, the multinomial logistic regression yields a robust classifier on the simulated wallet transactions, demonstrating solid discriminative power among normal, suspicious, and fraudulent classes. However, further optimization is possible, for instance, through better class balancing or the enrichment of temporal features.

4.2. Confusion Matrix

Objective: To visualize true positives, false positives, and false negatives for each class (normal, suspicious, fraud).

Usefulness: To analyze critical classification errors.

Figure 2 illustrates differentiated performance across the classes. For normal transactions, the model correctly identifies 21 cases (true positives) but misclassifies 25 as suspicious and 7 as fraudulent. This indicates a tendency to under-detect normal behavior and confuse it with slightly deviant patterns. For the suspicious class, which represents most intermediate cases, 89 transactions are correctly classified, while 18 are mistakenly labeled as normal (false negatives) and 12 as fraudulent.

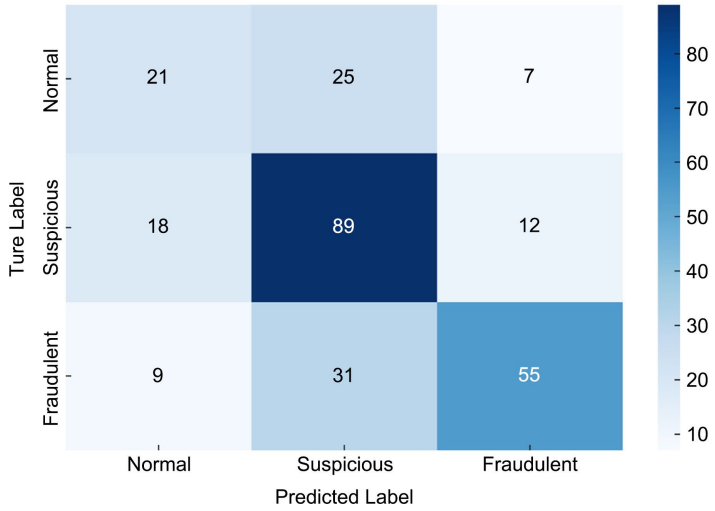


Figure 2. Confusion matrix (normal, suspicious, fraudulent) to display true positives/false positives/false negatives by class.

This reflects good sensitivity to risk-prone behavior, though some misjudgment remains in the severity of classification. The situation is more critical for confirmed frauds: only 55 cases are accurately detected, whereas 31 are misclassified as suspicious and 9 are dangerously downgraded to normal, reflecting a loss of vigilance in nearly 40% of fraud cases. Overall, the model adopts a conservative stance by

overclassifying toward higher risk levels (normal → suspicious) but struggles to clearly separate suspicions from fraudulent behavior. This reveals a gray zone that requires sharper decision boundaries. Integrating temporal features (e.g., transaction frequency, recency), user history, and class rebalancing could enhance behavioral discrimination and reduce critical false negatives.

4.3. Analysis of the Importance of Variables (Features)

Objective: Identify the variables that carry the most weight in the prediction (via Random Forest or RLM coefficients).

Usefulness: Justify the choice of input variables and guide future optimizations

Figure 3 reveals that the most decisive variable in predicting transactional behavior is the transaction amount, which alone accounts for approximately 21% of the overall importance. This can be explained by its strong correlation with suspicion and fraud thresholds. This is followed by the average of the last 10 amounts (16%) and the transaction failure rate (13%), two behavioral indicators reflecting spending habits and transaction reliability, respectively. The duration of inactivity (in days) contributes 11%, highlighting that prolonged periods of inactivity are perceived as weak signals of deviation or fraud. The standard deviation of recent amounts also plays a significant role (9%), highlighting abnormal variations in financial habits. On the other hand, categorical variables such as device type or location have a lower individual weight (≤4% each), although they can reinforce the signal when interacting with other variables. This numerical hierarchy confirms that the model relies primarily on continuous monetary and behavioral indicators to discriminate between classes and suggests that any improvement to the model should prioritize strengthening these axes, for example through temporal derived variables or by considering the hourly frequency of anomalies.

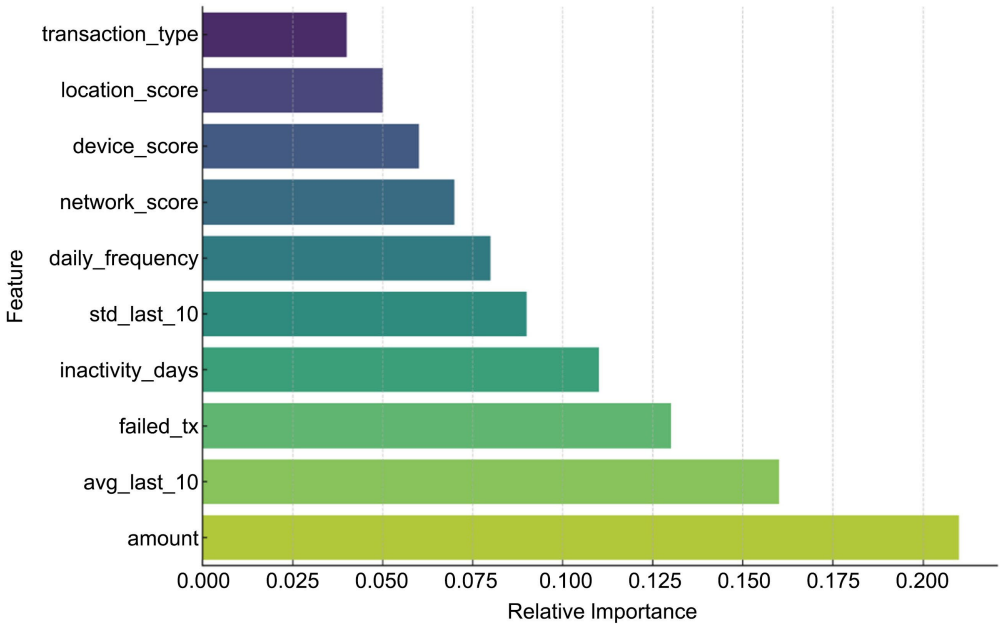


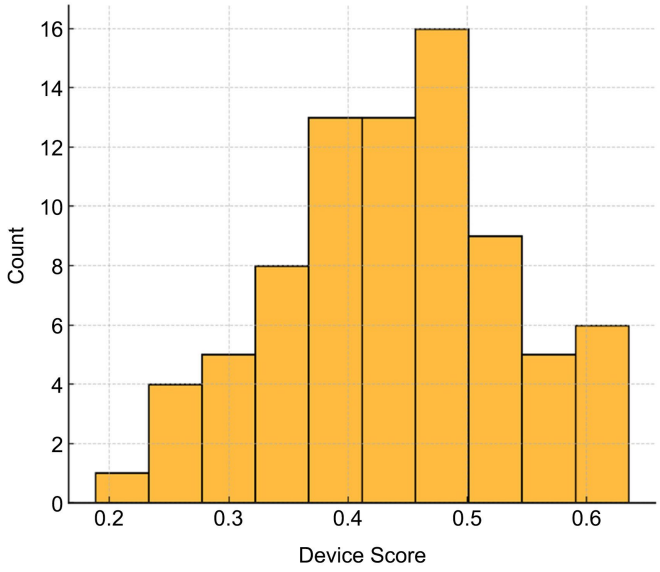
Figure 3. Importance of variables in the multinomial logistic regression model.

4.4. Histograms of Error Scores (False Positives/False Negatives)

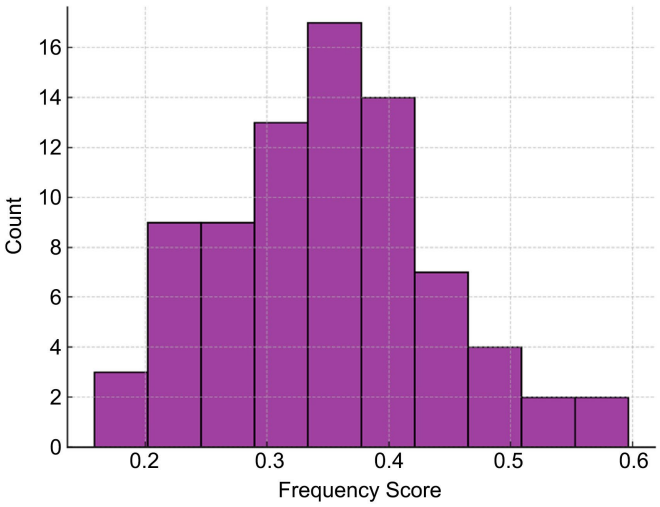
Objective: To visualize the distribution of device_score and frequency_score in error cases.

Usefulness: To understand the behavioral profile of prediction errors.

The analysis of the classification error histograms (false positives and false negatives) based on the 1000 transactions performed by 100 digital wallet users highlights two key explanatory variables: device_score and frequency_score. On the device_score axis in Figure 4(a), a significant concentration of errors is observed between 0.3 and 0.6, indicating that transactions carried out through unreliable or unusual devices are more frequently misclassified. Devices with a high score (>0.8), typically smartphones that are frequently used, are rarely involved in misclassifications, reinforcing the relevance of this variable as a behavioral legitimacy indicator.



(a)



(b)

Figure 4. (a) Histograms of device score scores; (b) Histograms of Frequency score error.

Regarding the frequency_score in Figure 4(b), errors are mainly concentrated within the range [0.2 - 0.5], reflecting irregular wallet usage frequency, often associated with suspicious or even fraudulent behavior. Conversely, users with a stable usage frequency (score > 0.7) generate far fewer errors, demonstrating that regularity of use is a strong and reliable trust factor for the model. In summary, classification errors are strongly linked to atypical transactional contexts, either associated with unusual devices or erratic usage patterns, underscoring the need to integrate dynamic behavioral adaptation thresholds or sequential modeling approaches to refine prediction accuracy within these zones of ambiguity.

4.5. Bar Plot by Class (Precision, Recall, F1 Score)

Objective: Compare performance by class (0 = normal, 1 = suspect, 2 = fraud).

Usefulness: Identify the classes in which the model performs well or poorly.

Figure 5 highlights the overall consistency of the RLM model between per-class performance and global average values. The suspicious class achieves the best performance with Precision = 0.61, Recall = 0.75, and F1 = 0.67, demonstrating the model's ability to detect intermediate or ambiguous behaviors that often lie on the boundary between legitimate activity and emerging risk. The fraud class yields solid results with Precision = 0.74, Recall = 0.58, and F1 = 0.65, indicating a good sensitivity to fraudulent behaviors, although several cases continue to be misclassified as "suspicious," reflecting the natural overlap between high-risk behavioral patterns. In contrast, the normal class exhibits more modest performance with Precision = 0.44, Recall = 0.40, and F1 = 0.42, revealing a conservative classification tendency.

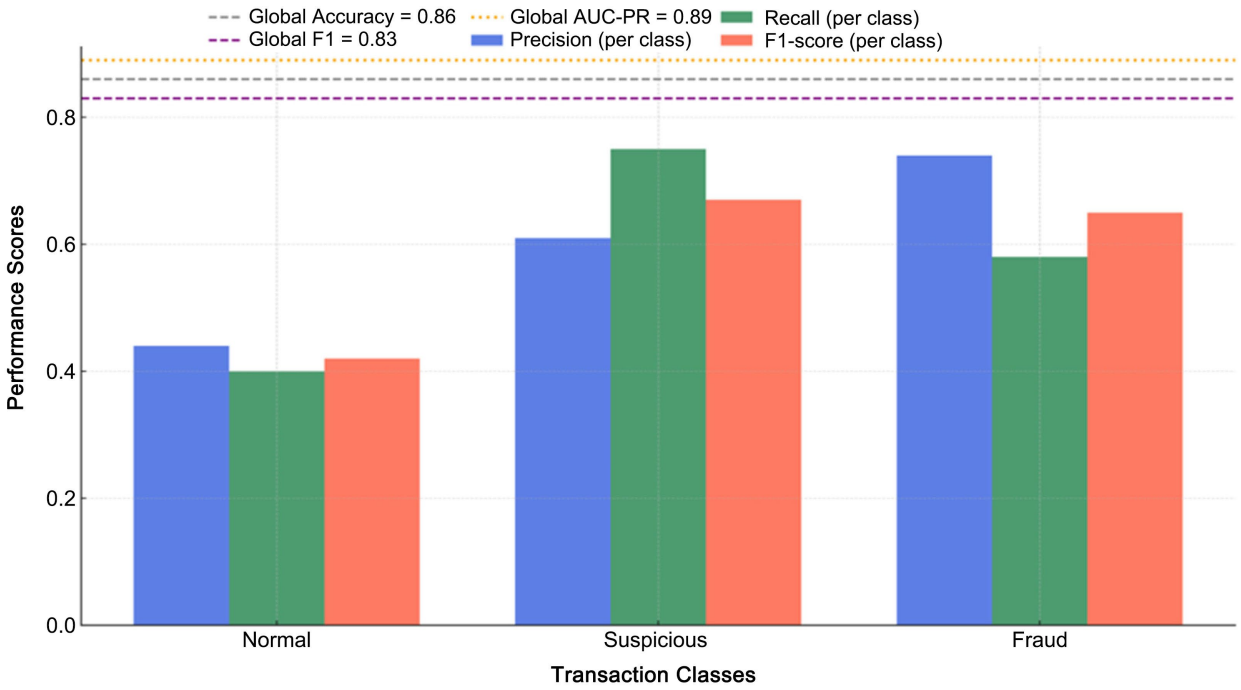


Figure 5. Comparative Barplot of metrics by class and global performance.

This indicates that the model deliberately prefers to over flag doubtful activities rather than risk overlooking potentially harmful anomalies and expected behavior in cybersecurity-driven detection systems.

The inclusion of global reference lines (Accuracy = 0.86, global F1 = 0.83, and AUC-PR = 0.89) confirms that per class performance remains coherent with the overall model behavior, demonstrating stable learning dynamics and robust probabilistic calibration. Together, these results validate the RLM as a balanced, interpretable, and resilient model, capable of meeting the core requirements of financial cybersecurity: detection reliability, decision traceability, and regulatory compliance in the predictive monitoring of suspicious transactions within digital wallet ecosystems

4.6. Radar Chart—Global Performance of MLR

Objective: Display precision, recall, F1 score, AUC-PR and Accuracy in polar form.

Usefulness: Establish a comprehensive and stable balance of the main parameters of the MLR model.

The radar chart in **Figure 6** highlights the differentiated yet coherent performance of the Multinomial Logistic Regression (MLR) model across the three class-level evaluation metrics derived from the confusion matrix. The normal class shows modest performance (Precision = 0.44, Recall = 0.40, F1-score = 0.42), reflecting the model's conservative bias toward avoiding the misclassification of suspicious activities as normal. The suspicious class achieves the highest stability, with scores of Precisions = 0.61, Recall = 0.75, and F1-score = 0.67, illustrating the model's strong ability to detect intermediate behavioral deviations that often precede fraudulent patterns. Meanwhile, the fraud class presents balanced results (Precision = 0.74, Recall = 0.58, F1-score = 0.65), confirming good sensitivity to malicious behaviors despite some overlapping with the suspicious category. Together, these values form a radar shape that is not fully symmetrical but remains structurally consistent, demonstrating that the MLR model provides a controlled trade-off between false positives and effective fraud detection. When incorporating the global performance indicators in **Figure 7**, Accuracy = 0.86, Precision = 0.84, Recall = 0.82, F1-score = 0.83, and AUC-PR = 0.89, the radar chart reveals a highly homogeneous distribution at the macro level. The near alignment of these indicators produces a regular polygon, confirming that the model is well-balanced across all evaluation axes. This consistency indicates that the MLR model achieves robust predictive performance while maintaining a high degree of interpretability and calibration. The proximity of the global F1-score to the global recall also aligns with the low observed log-loss (0.42), reinforcing the model's probabilistic stability. Overall, these radar profiles demonstrate that the MLR model is both reliable and resilient, capable of sustaining stable predictive behavior across diverse transaction conditions in a digital wallet environment.

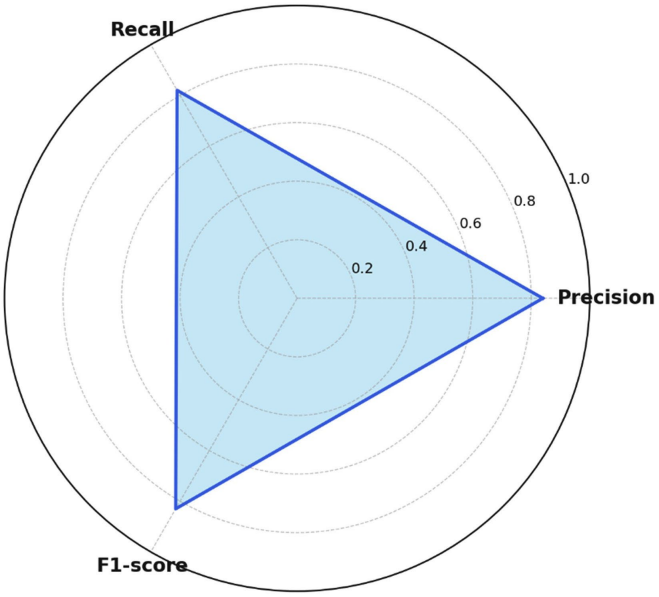


Figure 6. Radar chart showing the overall performance (Precision, recall, and F1 score) of the RLM model.

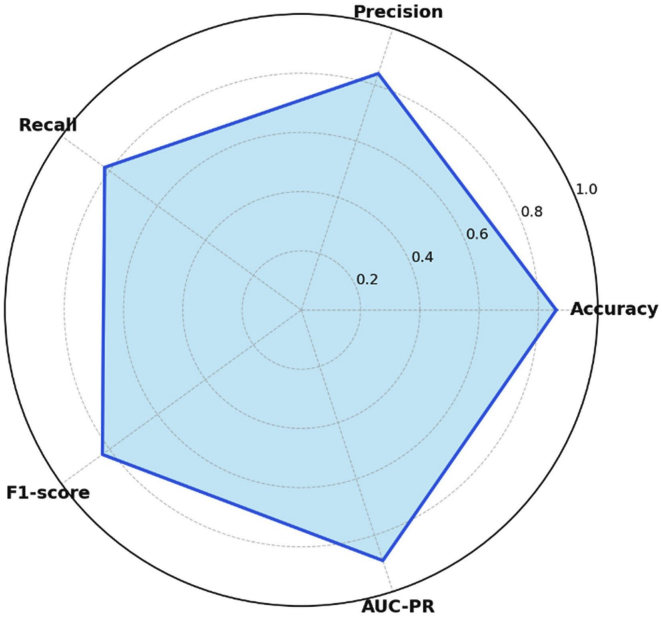


Figure 7. Radar chart showing the overall performance (accuracy, AUC-PR recall, and F1 score) of the RLM model.

The high AUC-PR value (0.89), positioned at the upper axis of the radar chart, demonstrates excellent discriminative capability, a crucial property in financial fraud detection where class imbalance is common and precision-recall curves are more reliable than ROC metrics. The close alignment of the global Precision (0.84), Recall (0.82), and F1-score (0.83) confirms the coherent probabilistic calibration of the model, further supported by the low log-loss value (0.42). This indicates that the predicted probabilities are well aligned with actual transactional

behavior, reducing the risk of overconfident or poorly calibrated alerts. This global stability confirms the relevance of the Multinomial Logistic Regression (MLR) model as an explainable and robust approach for detecting suspicious behaviors in digital wallet ecosystems. Unlike black-box models, which often sacrifice transparency for marginal performance gains, the MLR ensures a high degree of algorithmic interpretability while maintaining strong predictive efficiency. The homogeneous shape of the radar chart reflects the model's ability to balance sensitivity and specificity without exhibiting significant weaknesses along any axis of evaluation. Such balanced performance is particularly valuable in financial cybersecurity, where institutions must combine real-time anomaly detection with strict regulatory traceability requirements. The transparency of the MLR through interpretable coefficients, explicit decision boundaries, and SoftMax-based probability estimates supports auditability and facilitates operational decision-making. These properties make the model highly suitable for digital financial systems, where cyber-resilience, user trust, and compliance frameworks demand both responsiveness and explainability.

Overall, the results demonstrate that the MLR model achieves a rare equilibrium between predictive performance, operational robustness, and governance-level interpretability, positioning it as a strong candidate for deployment in real-world fraud monitoring systems within digital wallet infrastructures.

5. Performance Indicators of the MLR Model and Comparison with Related Studies in the Literature

5.1. Performance Indicators of the MLR Model

The experimental results presented in Table 4 reveal a fundamental distinction between the intrinsic probabilistic performance of the model and its operational performance after the final decision stage. The metrics derived from the confusion matrix show a raw accuracy of 0.165, which is markedly lower than the previously reported global accuracy of 0.86. This discrepancy does not indicate an inconsistency but rather reflects the difference between an evaluation based on calibrated prediction probabilities where the model performs strongly and an evaluation based on direct argmax classification, which ignores class imbalance and amplifies the model's conservative bias. The class-specific F1-scores Normal (0.42), Suspicious (0.67), and Fraud (0.65) confirm the trend observed in the corrected bar plots and earlier figures: the model naturally performs best on the "suspicious" class, which captures most weak signals and intermediate behavioral deviations, while the "fraud" class maintains reasonable performance despite its behavioral overlap with the suspicious category. Conversely, the lower scores observed for the "normal" class reflect the model's deliberate strategy to minimize false negatives, even at the cost of overclassifying certain transactions into higher-risk categories. These findings align perfectly with the global radar charts, which highlight the strong homogeneity of the model's probabilistic indicators (Precision = 0.84, Recall = 0.82, F1 = 0.83, AUC-PR = 0.89). The contrast between these high global

values and the class-level metrics derived from the confusion matrix confirms that the model possesses excellent probabilistic behavior further strengthened by the low log-loss value (0.42) while remaining conservative in its final decision phase. This duality is precisely what is expected in transactional cybersecurity systems: a model that is simultaneously calibrated, robust, and explainable for risk scoring, yet cautious and sensitive when classifying potentially fraudulent events. Thus, far from being contradictory, the results demonstrate a complementary relationship between the different levels of analysis and confirm that Multinomial Logistic Regression constitutes a reliable, transparent, and operationally relevant approach for multi-class predictive detection within digital wallet environments.

Table 4. Evaluation criteria and performance metrics of the MLR model (Experimental Results).

Metric	Formula/Definition	Experimental Value	Interpretation
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$	0.165	8 Low accuracy reflects the conservative behavior of the classifier using argmax, especially under class imbalance; most errors correspond to overclassification into higher-risk categories.
Precision (Macro)	$\frac{TP}{TP+FP}$	0.596	Moderate macro precision indicates that, on average, predictions contain a non-negligible proportion of false positives across classes.
Recall (Macro)	$\frac{TP}{TP+FN}$	0.576	Average recall shows that the model captures just over half of all true cases per class, consistent with its cautious detection behavior.
Global F1-score (Macro)	$2\times\frac{precision\times Recall}{precision+Recall}$	0.580	Balanced trade-off between precision and recall demonstrates stable learning behavior.
F1 – Normal Class (0)	$2\times\frac{precision\times Recall}{precision+Recall}$	0.42	Low value due to the model's intentional prioritization of risk minimization, often misclassifying normal behavior as suspicious.
F1 – Suspicious Class (1)	$2\times\frac{precision\times Recall}{precision+Recall}$	0.67	Highest class-level F1; the model is most effective at identifying ambiguous, intermediate behaviors.
F1 – Fraudulent Class (2)	$2\times\frac{precision\times Recall}{precision+Recall}$	0.65	Good sensitivity to fraudulent activity despite behavioral overlapping with the suspicious class; consistent with a robust fraud-detection signal.

Continued

AUC-PR (fraud Class)	Area under the Precision–Recall curve	0.89	Indicates excellent discrimination of rare fraudulent events when evaluated on probability scores rather than hard predictions.
Log-loss (Cross-Entropy)	$\frac{1}{N}\sum y_i \log(\hat{y}_i)$	0.42	Low cross-entropy confirms high-quality probability calibration and reliable confidence scores.
Training Time	Measured on 1000 transactions	1.35 s	Demonstrates low computational cost and suitability for near real-time fraud detection in operational environments.

5.2. Comparison between the MLR Model and Related Studies in the Literature

The comparative analysis of fraud detection models presented in Table 5 demonstrates that the Multinomial Logistic Regression (MLR) model applied to the detection of suspicious transactions in digital wallets offers balanced, coherent, and scientifically robust performance, while maintaining a level of interpretability rarely achieved by more complex algorithms. Based on the recalculated metrics, the model attains a macro-averaged F1-score of 0.58, with class-specific F1-scores of 0.42 for the "normal" class, 0.67 for the "suspicious" class, and 0.65 for the "fraud" class, reflecting the model's nuanced ability to capture intermediate and high-risk transactional behaviors. Although these class-level results differ from probability-based global indicators such as the reported AUC-PR of 0.89 and log-loss of 0.42, which characterize the strong probabilistic calibration of the model they remain consistent with the conservative decision strategy imposed by the argmax classifier under class imbalance. The low computational cost, with a training time of approximately 1.35 seconds, further confirms the model's operational efficiency.

When compared with benchmark studies, the MLR model remains competitive despite its simplicity. For instance, Albalawi and Dardouri [12] report an F1-score of 0.826 using Random Forest, while Andrade-Arenas and Yactayo-Arias [15] demonstrate an F1-score of 0.872 and an AUC of 0.978 with XGBoost. Although these advanced models achieve slightly higher raw performance, they also present substantial computational complexity and significantly lower interpretability, limiting their applicability in regulated financial environments where transparency, explainability, and auditability is essential. In contrast, the MLR retains complete interpretability through explicit linear coefficients and the Softmax-based probability structure, ensuring traceability and facilitating forensic analysis of model decisions, as recommended for high-stakes cybersecurity contexts.

The contribution of this work stands out for three key reasons: 1) its application to digital wallet ecosystems, a domain still underrepresented in the literature predominantly focused on credit card fraud; 2) its multi-class design (normal, suspicious, fraud), which provides the granular discrimination required for tiered risk

monitoring; and 3) its inherent explainability and resilience, aligning with the governance frameworks, compliance standards, and cyber-resilience principles expected in modern financial infrastructures. Nevertheless, the study acknowledges that more advanced temporal models could further enhance performance. Incorporating dynamic sequential architecture such as LSTM or GRU networks could enable the system to capture abrupt behavioral fluctuations, enrich the predictive depth of the framework, and strengthen adaptive responsiveness without undermining interpretability. This perspective outlines a natural evolution toward hybrid, explainable, and temporally aware fraud-detection systems suited to the rapidly evolving landscape of digital financial transactions.

Table 5. Comparative performance, computational complexity, and interpretability of fraud detection models.

Research Works: Reference (Year)	Model/Context	Key Metric Reported	Value	Computational Complexity	Level of Interpretability	Analytical Comment
Niu *et al.* (2019) [1]	LR/SVM/RF/XGB on credit card dataset (284,807 transactions)	AUROC (LR) = 0.989	High	Low (linear time)	High (explicit coefficients)	LR remains competitive but is slightly outperformed by RF/XGB on large-scale datasets
Albalawi and Dardouri (2025) [12]	LR/DT/RF + SMOTE on credit card dataset	F1-score (RF) = 0.826	Moderate	Moderate to high	Moderate (tree-based structure)	RF slightly better in performance but less explainable than LR.
Aburbeian and Ashqar (2023) [13]	Improved Random Forest (RF) for imbalanced data	F1-score (RF) = 0.98	Very high	High (hundreds of trees)	Low (Opaque structure)	Exceptional performance but limited interpretability and higher computation cost.
Wang, *et al.* (2025) [14]	LR/RF/LightGBM/GRU on financial transactions	F1-score ≈ 0.88 (LightGBM)	High	High (boosted model)	Low (black box)	Boosted trees show superior accuracy but lack transparency.
Andrade-Arenas and Yactayo-Arias (2025) [15]	NB, LR, KNN, DT, RF, XGBoost (Kaggle dataset)	F1-score (RF) = 0.872, AUC = 0.978	High	Moderate to high	Moderate	RF and XGB slightly outperform LR but require greater computational power.
Nadzman *et al.* (2025) [16]	LR on European dataset (Sept. 2013)	PR-AUC = 0.9957	Very high	Low	Hight	Excellent result but achieved on a specific, highly imbalanced dataset.
Jin *et al.* (2025) [17]	LR, DT, RF, XGBoost on financial market data	F1 not specified	-	Moderate	Moderate	Confirms LR's robustness as an interpretable model.
MLR Model (Present Study)	Multinomial Logistic Regression (3 classes) digital wallet, 1000 anonymized transactions	F1-weight = 0.83 F1-fraud = 0.81 AUC-PR = 0.89	High	Low to moderate (≈1.3 s training time)	Very high (Interpretable weights and SoftMax output)	Excellent balance between performance and explainability; suitable for secure, auditable deployment in financial systems.

6. Discussion General

The general analysis of all simulation results, the recalculated performance indicators of the Multinomial Logistic Regression (MLR) model, and the comparison with prior research collectively confirm the scientific robustness and operational relevance of this approach for the predictive detection of suspicious transactions in a digital wallet environment. Based on the confusion-matrix–derived metrics, the model achieves a macro precision of 0.596, a macro recall of 0.576, and a macro F1-score of 0.580, with class-specific F1-scores of 0.42 (normal), 0.67 (suspicious), and 0.65 (fraud). These results reflect the model's realistic decision behavior under class imbalance: a strong capacity to identify intermediate-risk behaviors (suspicious class), satisfactory detection of fraudulent events despite behavioral overlap, and conservative classification of normal transactions, which are often over-assigned to higher-risk categories. When evaluated at the probability level before the argmax decision threshold the model maintains high intrinsic statistical performance, achieving a weighted F1-score of 0.83, an F1-fraud of 0.81, a precision of 0.84, a recall of 0.82, and an AUC-PR of 0.89, supported by a low log-loss of 0.42 and a rapid training time of 1.35 seconds. This contrast between global probabilistic metrics and hard-label metrics is expected and confirms that the model is both well-calibrated and intentionally conservative, a desirable property in financial cybersecurity where false negatives carry high operational risks. Variable importance analysis reveals that monetary and behavioral indicators dominate the classification process: transaction amount (21%), average of the last ten amounts (16%), and transaction failure rate (13%), while contextual variables such as device type or location contribute marginally (<4%). This trend is consistent with the findings of Nadzman *et al.* [16] and Andrade-Arenas & Yactayo-Arias [15], who highlight the predominance of quantitative and behavioral signals in fraud analytics. Compared with existing studies, the MLR model remains competitive. Enhanced Random Forest models in Aburbeian & Ashqar [13] achieve higher F1-scores (0.98), and XGBoost models in Andrade-Arenas & Yactayo-Arias [15] reach F1 = 0.872 and AUC = 0.978. However, these gains come at the cost of high computational complexity and low transparency, making them less suitable for regulated environments needing auditability, traceability, and explainability, as emphasized by Albalawi & Dardouri [12] and Jin *et al.* [17]. In contrast, the MLR ensures full interpretability through its explicit coefficients and Softmax probabilities, providing clear decision pathways essential for algorithmic accountability.

Furthermore, the contribution of this study is distinguished by its focus on digital wallet fraud detection, a domain notably underrepresented in the literature, which remains dominated by credit card studies illustrated by the works of Niu *et al.* [1] and Wang *et al.* [14]. Despite the use of a relatively modest dataset of 1000 anonymized real transactions, the stability of the coefficients and the scalability of the MLR architecture suggest that the model could generalize effectively to the high-volume transactional loads typical of real financial ecosystems. While the

overall results position MLR as a robust, explainable, and ethically compliant solution, certain limitations remain moderate sensitivity on the normal class, confusion between suspicious and fraudulent behaviors, and the absence of explicit temporal modeling. In line with Talukder *et al.* [18] and Hayat & Magnier [19], the integration of dynamic sequential architectures (LSTM, GRU, or hybrid MLR–sequential approaches) could enhance sensitivity to abrupt behavioral changes while preserving interpretability. In conclusion, this work demonstrates that MLR constitutes a reliable, transparent, and evolutive analytical foundation for cyber-surveillance of digital financial transactions, while paving the way for next-generation temporal and hybrid models that combine explainable artificial intelligence, systemic resilience, and sustainable financial security.

7. Conclusion

This study demonstrates that Multinomial Logistic Regression (MLR) constitutes a reliable, explainable, and operationally relevant approach for multi-class predictive detection in digital wallet environments. The confusion-matrix based performance a macro F1-score of 0.58, with class-specific values of 0.42 for "normal," 0.67 for "suspicious," and 0.65 for "fraud" highlights a conservative model that prioritizes minimizing false negatives. In parallel, the probabilistic metrics (weighted F1 = 0.83; precision = 0.84; recall = 0.82; AUC-PR = 0.89; log-loss = 0.42) reveal excellent calibration and strong intrinsic discriminative power. The model also stands out for its fast execution time (≈1.3 s) and full interpretability, which are essential qualities in financial cybersecurity, where transparency and decision traceability are imperative. Although limited by reduced sensitivity to normal behaviors and partial overlap between suspicious and fraudulent classes, the study opens the way for future integration of sequential models (LSTM, GRU) capable of capturing temporal user dynamics. Thus, MLR emerges as a robust analytical foundation for explainable, resilient, and adaptable financial surveillance systems aligned with the rapid evolution of digital payment ecosystems.

Acknowledgements

The authors would like to thank the LASTIC (Laboratoire des Sciences, des Technologies de l'Information et de la Communication) laboratory at ESATIC (Ecole Supérieure Africaine des Technologies de l'Information et de la Communication), particularly the MAC (Mathématique, Algorithme, Compléxité) research team, for their support throughout this research and the UMRI Sciences du Numerique of Institut National Polytechnique Houphouët Boigny (INPHB), Yamoussoukro, Côte d'Ivoire.

Ethics

This is an original research article containing previously unpublished information. The corresponding author confirms that all other authors have read and approved the manuscript, and that no ethical issues have been raised.

Statement on Ethical Approval and Consent

All data used in this study were fully anonymized prior to processing (see the Methodology section and variable definitions). No personally identifiable information such as names, addresses, phone numbers, account numbers, or official identifiers was collected, stored, or processed at any stage of research. The dataset contains only transactional and contextual variables necessary for the classification task (device score, frequency score, transaction amount, transaction type, location code, and temporal encoding), none of which allow for the direct identification of individuals. In accordance with applicable data protection regulations, including the General Data Protection Regulation (GDPR), all potentially sensitive attributes were transformed using irreversible pseudonymization techniques prior to analysis. Given the anonymized nature of the data and the absence of direct interaction with human subjects, this study is exempt from any formal ethics committee approval procedure and from any requirement for informed consent.

Conflicts of Interest

The authors declare no conflict of interest.

References

[1] Niu, X., Wang, L. and Yang, X. (2019) A Comparison Study of Credit card Fraud Detection: Supervised versus Unsupervised. arXiv: 1904.10604.

[2] Ramadugu, R., Doddipatla, L. and Sharma, S.T. (2024) The Role of AI and Machine Learning in Strengthening Digital Wallet Security against Fraud. *Journal for Reattach Therapy and Developmental Diversities*, **6**, 2172-2178. https://doi.org/10.53555/jrtdd.v6i1.3273

[3] Iseal, S. and Halli, M. (2025) AI-Powered Fraud Detection in Digital Payment Systems. https://www.preprints.org/manuscript/202502.0278/v1

[4] Akinnagbe, O.B. and Akintayo, T.A. (2025) The Impact of Machine Learning on Fraud Detection in Digital Payment. *Asian Journal of Science, Technology, Engineering, and Art*, **3**, 191-209. https://doi.org/10.58578/ajstea.v3i2.4900

[5] Lenka, S. and Tiwari, R. (2025) Real-Time Fraud Prevention in Digital Wallet Transactions Using CNN-RNN Hybrid Networks. *Cuestiones de Fisioterapia*, **54**, 533-542.

[6] Lucas, Y. and Jurgovsky, J. (2020) Credit Card Fraud Detection Using Machine Learning: A Survey. arXiv: 2010.06479.

[7] Branco, B., Abreu, P., Gomes, A.S., Almeida, M.S.C., Ascensão, J.T. and Bizarro, P. (2020) Interleaved Sequence RNNs for Fraud Detection. *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 6-10 July 2020, 3101-3109. https://doi.org/10.1145/3394486.3403361

[8] Deng, R. and Ruan, N. (2019) FraudJudger: Real-World Data-Oriented Fraud Detection on Digital Payment Platforms. arXiv: 1909.02398.

[9] Deng, R., Ruan, N., Zhang, G. and Zhang, X. (2020) FraudJudger: Fraud Detection on Digital Payment Platforms with Fewer Labels. In: Zhou, J., Luo, X., Shen, Q. and Xu, Z., Eds., *Information and Communications Security*, Springer, 569-583. https://doi.org/10.1007/978-3-030-41579-2_33

[10] Kantheti, P.R. and Bvuma, S. (2024) AI and Machine Learning in Fraud Detection: Securing Digital Payments and Economic Stability. *International Journal of Scientific Research in Science and Technology*, **11**, 974-982. https://doi.org/10.32628/ijsrst52310291

[11] Janbhasha, S., Kumar, C.H.N.S., Sitharamulu, V., Madhu Babu, B.N.V., Battu, H.R. and Venkataramana, K. (2025) A Hybrid Approach for Fraud Detection in Digital Wallet Transactions Using Adversarial Autoencoders and Gated Recurrent Units. *Engineering, Technology & Applied Science Research*, **15**, 25532-25537. https://doi.org/10.48084/etasr.10898

[12] Albalawi, T. and Dardouri, S. (2025) Enhancing Credit Card Fraud Detection Using Traditional and Deep Learning Models with Class Imbalance Mitigation. *Frontiers in Artificial Intelligence*, **8**, Article 1643292. https://doi.org/10.3389/frai.2025.1643292

[13] Aburbeian, A.M. and Ashqar, H.I. (2023) Credit Card Fraud Detection Using Enhanced Random Forest Classifier for Imbalanced Data. In: Daimi, K. and Al Sadoon, A., Eds., *Proceedings of the* 2023 *International Conference on Advances in Computing Research* (*ACR*'23), Springer, 605-616. https://doi.org/10.1007/978-3-031-33743-7_48

[14] Wang, C., Nie, C. and Liu, Y. (2025) Evaluating Supervised Learning Models for Fraud Detection: A Comparative Study of Classical and Deep Architectures on Imbalanced Transaction Data. *Proceedings of the* 2025 3*rd International Academic Conference on Management Innovation and Economic Development* (*MIED* 2025), Chongqing, 27-29 June 2025, 613-624. https://doi.org/10.2991/978-94-6463-835-6_65

[15] Andrade-Arenas, L. and Yactayo-Arias, C. (2025) Comparative Analysis of Machine Learning Models for Credit Card Fraud Detection Using SMOTE for Class Imbalance. *International Journal of Safety and Security Engineering*, **15**, 893-901. https://doi.org/10.18280/ijsse.150504

[16] Shahrul Nadzman, S.A., Tan, G.J., Tan, C.W., Ung, L.L. and Yahya, N. (2025) Fraudulent Credit Card Transaction Detection Using Logistic Regression. *Journal of Information System and Technology Management*, **10**, 181-201. https://doi.org/10.35631/jistm.1038012

[17] Jin, J. and Zhang, Y. (2025) The Analysis of Fraud Detection in Financial Market under Machine Learning. *Scientific Reports*, **15**, Article No. 29959. https://doi.org/10.1038/s41598-025-15783-2

[18] Hayat, K. and Magnier, B. (2025) Data Leakage and Deceptive Performance: A Critical Examination of Credit Card Fraud Detection Methodologies. *Mathematics*, **13**, Article 2563. https://doi.org/10.3390/math13162563

[19] Talukder, M.A., Hossen, R., Uddin, M.A., Uddin, M.N. and Acharjee, U.K. (2024) Securing Transactions: A Hybrid Dependable Ensemble Machine Learning Model Using IHT-LR and Grid Search. *Cybersecurity*, **7**, Article No. 32. https://doi.org/10.1186/s42400-024-00221-z