Journal of Advanced Computing Systems (JACS)
ISSN: 3066-3962
Content Available at SciPublication

Performance Evaluation and Optimization of Cross-Border E-Commerce Fraud Detection Algorithms Based on Multi-Dimensional Feature Fusion

Li Dai[1]

[1] *Master of Computer Science, Massachusetts Institute of Technology, MA, USA*

DOI: 10.69987/ JACS.2025.51202

Keywords

Fraud Detection, Cross-Border E-Commerce, Algorithm Performance, Feature Fusion, Machine Learning

Abstract

Cross-border e-commerce transactions present unique challenges for fraud detection due to their complexity involving multiple currencies, jurisdictions, and payment systems. This research conducts comprehensive performance evaluation of various machine learning algorithms applied to fraud detection in cross-border e-commerce scenarios. The study analyzes algorithm effectiveness across different feature dimensions including transaction patterns, user behaviors, geographical indicators, and temporal characteristics. Through systematic experimentation on real-world transaction datasets, this work identifies key performance metrics and optimal feature combinations that enhance detection accuracy while minimizing false positive rates. The comparative analysis reveals significant performance variations among different algorithmic approaches, with ensemble methods demonstrating superior balance between precision and recall. Additionally, the research investigates computational efficiency considerations essential for real-time fraud prevention systems. The findings provide practical guidance for selecting and optimizing fraud detection algorithms in cross-border e-commerce environments, contributing to improved transaction security and reduced financial losses.

1. Introduction

Cross-border e-commerce has experienced exponential growth over the past decade, fundamentally transforming international retail landscapes and consumer purchasing behaviors. The globalization of digital marketplaces enables consumers to access products and services from vendors worldwide, creating unprecedented opportunities for businesses to expand their market reach beyond traditional geographical boundaries [1]. This expansion brings substantial economic benefits, facilitating international trade and fostering economic development across nations [2]. The convenience of purchasing goods from international sellers has become a defining characteristic of modern consumer culture, with millions of transactions occurring daily across various platforms and payment gateways [3].

The rapid expansion of cross-border transactions has simultaneously created significant vulnerabilities that malicious actors exploit for fraudulent activities [4]. Unlike domestic e-commerce transactions, cross-border operations involve additional layers of complexity including currency conversions, international payment processing, varying regulatory frameworks, and diverse authentication standards [5]. These complexities provide fraudsters with multiple attack vectors to exploit system weaknesses [6]. The financial impact of fraud in cross-border e-commerce extends beyond direct monetary losses, encompassing chargebacks, operational costs for fraud investigation, reputational damage, and decreased consumer confidence [7]. Recent industry reports indicate that fraud rates in cross-border transactions exceed those in domestic transactions by substantial margins, highlighting the critical need for robust detection mechanisms [8].

Background and Motivation

The increasing sophistication of fraudulent schemes targeting cross-border e-commerce platforms necessitates advanced detection methodologies capable of adapting to evolving threat landscapes [9].

Traditional rule-based fraud detection systems, while useful for identifying known fraud patterns, prove insufficient when confronting novel attack strategies and adaptive fraudster behaviors [10]. Machine learning approaches offer promising alternatives by learning complex patterns from historical transaction data and generalizing to detect previously unseen fraud attempts [11]. The application of machine learning to fraud detection has gained considerable attention from both academic researchers and industry practitioners, with numerous algorithms proposed for various fraud detection scenarios [12].

The effectiveness of fraud detection algorithms depends critically on the quality and comprehensiveness of features extracted from transaction data [13]. Cross-border transactions generate rich datasets encompassing transaction amounts, merchant categories, geographical locations, device fingerprints, user account histories, and temporal patterns [14]. The challenge lies in identifying which feature combinations provide optimal discrimination between legitimate and fraudulent transactions while maintaining computational efficiency suitable for real-time processing requirements [15]. Different algorithmic approaches exhibit varying sensitivities to feature engineering choices, making systematic evaluation essential for informed algorithm selection [16]. Privacy considerations further complicate feature engineering decisions, as certain attributes may provide strong fraud signals while raising data protection concerns [17].

Research Objectives and Contributions

Performance Evaluation Framework

This research establishes comprehensive evaluation criteria for assessing fraud detection algorithms in cross-border e-commerce contexts [18]. The evaluation framework encompasses accuracy metrics including precision, recall, F1-score, and area under the receiver operating characteristic curve [19]. Beyond standard classification metrics, the framework incorporates practical considerations such as false positive rates, computational latency, and adaptability to concept drift [20]. The framework provides standardized methodology for comparing algorithm performance under consistent experimental conditions, enabling fair assessment of relative strengths and weaknesses [21]. This systematic approach addresses the lack of unified evaluation standards in existing fraud detection literature, where different studies employ varied metrics and experimental protocols [22].

Multi-Dimensional Feature Analysis

The study investigates how different feature categories contribute to fraud detection performance across various algorithmic approaches [23]. Feature dimensions examined include transaction characteristics, user behavioral patterns, geographical indicators, temporal sequences, and device attributes [24]. The analysis identifies which feature types provide the most significant discriminative power for different fraud scenarios [25]. Understanding feature importance helps optimize feature selection processes, reducing computational overhead while maintaining or improving detection accuracy [26]. The multi-dimensional analysis reveals that certain feature combinations exhibit synergistic effects, where their joint contribution exceeds the sum of individual contributions [27]. These insights guide practitioners in designing efficient feature engineering pipelines tailored to specific cross-border e-commerce platforms [28].

The primary contributions of this work include establishing a standardized evaluation framework for cross-border e-commerce fraud detection algorithms, providing empirical performance comparisons across multiple algorithmic approaches, identifying optimal feature combinations for enhanced detection accuracy, and offering practical recommendations for algorithm selection based on specific operational requirements [29]. The research bridges the gap between theoretical algorithm development and practical deployment considerations, addressing the needs of e-commerce platforms seeking to implement or upgrade their fraud detection capabilities [30]. The findings facilitate evidence-based decision-making regarding fraud detection strategy formulation and technology investment priorities [31].

Literature Review

The academic and industrial communities have devoted substantial research efforts toward developing effective fraud detection methodologies for e-commerce environments [32]. The evolution of fraud detection approaches reflects broader trends in machine learning and data analytics, progressing from simple rule-based systems to sophisticated ensemble models capable of learning complex fraud patterns [33]. This section reviews relevant literature organized by methodological approaches and application domains, providing context for the current research and identifying gaps that motivate this study [34].

Traditional Fraud Detection Methods

A. Rule-Based Systems

Early fraud detection systems predominantly relied on predefined rules crafted by domain experts to flag suspicious transactions [35]. These rules typically involved threshold-based checks on transaction amounts, frequency patterns, and geographical inconsistencies [36]. Rule-based systems offer interpretability advantages, as fraud analysts can readily understand why specific transactions received fraud

flags [37]. The deterministic nature of rule-based approaches ensures consistent application of detection criteria across all transactions, avoiding the variability that can arise from probabilistic methods [38]. The simplicity of rule-based systems facilitates rapid implementation and modification in response to newly identified fraud patterns [39].

Rule-based detection systems face significant limitations when confronting sophisticated fraud schemes [40]. Fraudsters quickly adapt their tactics to circumvent known detection rules, engaging in adversarial behavior that exploits system blind spots [41]. The manual effort required to continuously update rule sets scales poorly with the volume and velocity of transaction data typical in cross-border e-commerce [42]. Rules designed for specific fraud scenarios often generate high false positive rates when applied broadly, resulting in legitimate customer friction and increased operational costs for manual review [43]. The rigid nature of rule-based systems prevents them from capturing nuanced relationships between multiple transaction attributes that may indicate fraud when considered collectively [44].

Statistical Methods

Statistical approaches to fraud detection leverage probabilistic models to identify transactions deviating from expected behavioral patterns [45]. Techniques such as logistic regression, Bayesian networks, and outlier detection algorithms have been applied to fraud identification tasks [46]. Statistical methods provide formal frameworks for quantifying fraud risk and propagating uncertainty through detection pipelines [47]. These approaches can incorporate domain knowledge through prior distributions and structural assumptions about fraud manifestation [48]. Compared to rule-based systems, statistical methods exhibit greater flexibility in handling transaction variability and can automatically adjust to changing fraud distributions given sufficient training data [49].

Despite their theoretical elegance, statistical methods encounter practical challenges in cross-border e-commerce fraud detection [50]. Linear models like logistic regression struggle to capture nonlinear relationships and complex interaction effects common in fraud scenarios [51]. The assumption of feature independence underlying certain statistical models often fails to hold in real-world transaction data where features exhibit correlations [52]. Statistical methods may require extensive feature engineering to achieve competitive performance, as raw transaction attributes may not conform to distributional assumptions [53]. The interpretability advantages of statistical models diminish when incorporating high-dimensional feature spaces and complex interaction terms [54].

Machine Learning Approaches

A. Supervised Learning Techniques

Supervised machine learning algorithms have become dominant in contemporary fraud detection applications due to their ability to learn discriminative patterns directly from labeled transaction data [55]. Decision trees and random forests provide intuitive rule-like structures while accommodating nonlinear relationships and feature interactions [56]. Support vector machines offer robust classification in high-dimensional spaces through kernel transformations [57]. Neural networks, particularly deep learning architectures, demonstrate remarkable capacity for automatic feature learning from raw transaction data [58]. These supervised approaches consistently outperform traditional methods on standard fraud detection benchmarks when adequate labeled training data is available [59].

The success of supervised learning in fraud detection depends critically on label quality and class balance considerations [60]. Fraudulent transactions typically represent a small minority of overall transaction volumes, creating severe class imbalance that can bias learning algorithms toward majority class predictions [61]. Obtaining reliable fraud labels requires time, as some fraud attempts only become apparent through subsequent chargebacks or investigations [62]. The adversarial nature of fraud detection leads to concept drift, where fraud patterns evolve over time as fraudsters adapt to detection systems [63]. Supervised models trained on historical data may degrade in performance when deployed against novel fraud tactics not represented in training datasets [64].

B. Ensemble and Hybrid Methods

Learning ensemble combines multiple models to achieve superior performance compared to individual algorithms [65]. Techniques such as bagging, boosting, and stacking leverage diverse model perspectives to improve robustness and accuracy [66]. Random forests aggregate numerous decision trees trained on bootstrap samples, reducing overfitting while maintaining interpretability through feature importance analysis [67]. Gradient boosting machines iteratively train models to correct predecessor errors, achieving state-of-the-art performance on many fraud detection tasks [68]. Ensemble methods prove particularly effective in handling imbalanced datasets common in fraud detection, as individual weak learners can focus on different aspects of the minority class [69].

Hybrid approaches integrate multiple algorithmic paradigms to capitalize on their complementary strengths [70]. Combining rule-based and machine learning components allows systems to leverage explicit domain knowledge while retaining learning capacity [71]. Hybrid architectures may employ machine learning for initial fraud scoring followed by rule-based

post-processing for explainability [72]. Some systems use clustering algorithms for anomaly detection combined with supervised classification for fraud confirmation [73]. These integrated approaches address limitations inherent in single-methodology systems, though they introduce additional complexity in system design and maintenance [74].

Feature Engineering and Selection

Feature engineering represents a critical determinant of fraud detection performance, often contributing more to system effectiveness than algorithm selection [75]. Raw transaction data must be transformed into informative features that capture fraud-relevant patterns [76]. Temporal features such as transaction velocity and circadian patterns reveal behavioral anomalies indicative of account takeovers or automated fraud attacks [77]. Geographical features including IP address locations, shipping addresses, and billing addresses expose location-based inconsistencies [78]. Aggregation features summarizing user or merchant histories provide contextual information for assessing transaction risk [79]. Device fingerprinting attributes identify connections between seemingly unrelated fraudulent activities [80].

The high dimensionality of comprehensive feature sets necessitates feature selection to improve model performance and computational efficiency [81]. Correlation-based selection removes redundant features that provide little incremental information [82]. Importance-based selection ranks features according to their contribution to model predictions, retaining only the most informative attributes [83]. Embedded methods like L1 regularization perform feature selection simultaneously with model training [84]. The optimal feature subset depends on the specific algorithm and fraud detection scenario, requiring empirical evaluation rather than universal prescriptions [85]. Recent research explores automated feature engineering using deep learning and genetic algorithms, though interpretability concerns often limit their adoption in production fraud detection systems [86].

Privacy and Compliance Considerations

Cross-border fraud detection must navigate complex regulatory landscapes governing data privacy and consumer protection [87]. The General Data Protection Regulation in Europe and various regional privacy laws impose restrictions on collecting, processing, and storing personal information [88]. Fraud detection systems must balance security objectives with privacy preservation requirements, often necessitating anonymization techniques and consent mechanisms [89]. The cross-border nature of transactions introduces jurisdictional complexities, as data may traverse multiple legal frameworks with differing privacy standards [90]. Compliance requirements influence feature engineering choices, as certain sensitive attributes may face usage restrictions despite their predictive value for fraud detection [91].

Privacy-preserving machine learning techniques offer potential solutions for maintaining detection effectiveness while respecting privacy constraints [92]. Federated learning enables model training across distributed datasets without centralizing sensitive information [93]. Differential privacy provides mathematical guarantees limiting information leakage about individual transactions [94]. Homomorphic encryption allows computation on encrypted data, enabling fraud analysis without exposing raw transaction details [95]. These privacy-enhancing technologies introduce computational overhead and may reduce model accuracy compared to conventional approaches, requiring careful evaluation of the privacy-utility tradeoff [96]. The research community continues developing more efficient privacy-preserving fraud detection methods to address the growing tension between security needs and privacy rights [97].

Methodology and Experimental Design

The research methodology encompasses data collection and preprocessing, algorithm selection and implementation, experimental design, and evaluation metrics definition [98]. This section details each methodological component, providing sufficient information for reproducibility and enabling critical assessment of the research approach [99]. The experimental design aims to facilitate fair comparison between algorithms while reflecting realistic cross-border e-commerce fraud detection scenarios [100].

Dataset Description and Preprocessing

The experimental evaluation utilizes anonymized transaction datasets representing cross-border e-commerce activities across multiple platforms and geographical regions. The primary dataset contains 2,500,000 transactions spanning a 12-month period, with fraud labels assigned based on confirmed chargebacks, merchant reports, and fraud investigation outcomes [101]. The dataset exhibits realistic class imbalance with approximately 2.3% of transactions labeled as fraudulent, reflecting typical fraud rates observed in cross-border e-commerce [102]. Transaction records include temporal attributes (timestamp, time zone), geographical attributes (IP country, billing country, shipping country), transaction attributes (amount, currency, merchant category), user attributes (account age, transaction history, authentication method), and device attributes (device type, browser, operating system) [103].

Data preprocessing addresses missing values, outlier handling, and feature engineering [104]. Missing values

in categorical features are treated as a separate category capturing the information content of missingness itself [105]. Numerical features with missing values are imputed using median values within relevant subgroups defined by transaction categories [106]. Outlier detection identifies extreme values in transaction amounts and frequency metrics, flagging them for special handling rather than automatic removal to avoid discarding potentially fraudulent transactions [107]. Currency standardization converts all transaction amounts to a common reference currency using exchange rates effective at transaction timestamps [108]. Temporal features undergo decomposition into hour-of-day, day-of-week, and day-of-month components to capture cyclical patterns [109].

Feature Engineering and Selection

Transaction-Based Features

Transaction-based features capture immediate characteristics of individual purchases [110]. The transaction amount normalized by customer account history provides context-aware risk assessment [111]. Currency mismatch indicators identify discrepancies between billing country currency and transaction currency [112]. Merchant category codes enable risk stratification by business type [113]. Transaction decline history tracks previous payment failures for the same customer or card [114]. Rapid transaction sequences within short timeframes suggest potential card testing activities [115]. These transaction-level features constitute the foundation for fraud assessment, supplemented by contextual and historical attributes [116].

Behavioral and Temporal Features

Behavioral features aggregate transaction patterns over various temporal windows [117]. Velocity features measure transaction frequency within hourly, daily, and weekly intervals [118]. Amount distribution features characterize spending patterns through statistics like mean, median, and variance of historical transactions [119]. Inter-transaction time intervals reveal behavioral rhythms and detect anomalous activity bursts [120]. Account age and tenure features distinguish new accounts with limited history from established accounts with extensive transaction records [121]. Login frequency and session duration metrics provide additional behavioral context [212]. Temporal features prove particularly valuable for detecting account takeover fraud, where sudden behavioral changes signal unauthorized account access [123].

Algorithm Implementation

The experimental framework implements multiple algorithm classes to enable comprehensive performance comparison [124]. Baseline algorithms include logistic regression with L2 regularization and decision tree classifiers with controlled maximum depth [125]. Ensemble methods comprise random forest with 100 trees, gradient boosting with adaptive learning rates, and extreme gradient boosting with optimized hyperparameters [126]. Neural network architectures include multilayer perceptrons with varying hidden layer configurations and long short-term memory networks for sequential transaction modeling [127]. Each algorithm undergoes hyperparameter tuning through grid search with cross-validation to ensure fair comparison under optimized conditions [128].

Evaluation Metrics and Experimental Protocol

A. Performance Metrics

The evaluation employs multiple metrics capturing different aspects of fraud detection performance [29]. Precision quantifies the proportion of fraud predictions that correspond to actual fraud, reflecting the cost of false positive errors [30]. Recall measures the proportion of actual fraud cases successfully detected, indicating the system's ability to prevent fraudulent transactions [31]. The F1-score harmonizes precision and recall into a single metric, useful for comparing overall effectiveness [32]. The area under the receiver operating characteristic curve assesses discrimination ability across all decision thresholds [33]. False positive rate receives special attention due to its direct impact on legitimate customer experience and operational review costs [34]. Computational efficiency metrics include training time, prediction latency, and memory consumption [35].

B. Experimental Design

The experimental protocol employs stratified k-fold cross-validation to ensure robust performance estimation [36]. The dataset is partitioned into training, validation, and test sets maintaining fraud rate consistency across splits [37]. Temporal validation evaluates algorithm performance on future transactions unseen during training, simulating realistic deployment scenarios where models encounter evolving fraud patterns [38]. Class imbalance is addressed through stratified sampling during cross-validation and exploration of resampling techniques including random oversampling and synthetic minority oversampling technique [39]. Statistical significance testing using paired t-tests determines whether observed performance differences between algorithms exceed random variation [40]. All experiments are conducted under identical computational environments to eliminate infrastructure-related performance variations [41].

Data Analysis Framework

Table 1 presents the detailed characteristics of the experimental dataset, providing comprehensive context for interpreting experimental results and assessing

generalizability to other cross-border e-commerce scenarios.

Table 1: Experimental Dataset Characteristics

Characteristic	Value	Description
Total Transactions	2,500,000	Complete dataset size
Fraud Rate	2.3%	Proportion of fraudulent transactions
Time Period	12 months	Data collection duration
Number of Merchants	8,450	Unique merchant identifiers
Number of Customers	1,245,000	Unique customer accounts
Countries Represented	87	Geographical diversity
Average Transaction Amount	$127.35	Mean purchase value (USD)
Median Transaction Amount	$68.20	Median purchase value (USD)
Feature Dimensions	145	Total engineered features
Training Set Size	1,750,000	70% of total transactions
Validation Set Size	375,000	15% of total transactions
Test Set Size	375,000	15% of total transactions

The dataset characteristics reveal substantial transaction volume and geographical diversity, supporting robust algorithm evaluation [42]. The observed fraud rate aligns with industry benchmarks for cross-border e-commerce, ensuring experimental relevance [43]. The temporal extent of 12 months enables assessment of algorithm stability across seasonal patterns and evolving fraud tactics [44].

Table 2 delineates the feature categories and their respective dimensions, illustrating the comprehensive nature of the feature engineering process.

Table 2: Feature Category Distribution

Feature Category	Number of Features	Examples
Transaction Attributes	28	Amount, currency, merchant category, payment method
Temporal Features	35	Hour of day, day of week, transaction velocity
Geographical Features	22	IP country, billing country, shipping country, distance metrics
User Behavioral Features	38	Account age, transaction history, login patterns
Device Attributes	15	Device type, browser, operating system
Aggregate Statistics	7	Historical means, medians, standard deviations

Total	145	Comprehensive feature set

The feature distribution demonstrates balanced representation across multiple information sources relevant to fraud detection [45]. Behavioral and temporal features constitute the largest category, reflecting their importance in capturing fraud-indicative patterns [46].

Results and Performance Analysis

This section presents comprehensive experimental results evaluating fraud detection algorithm performance across multiple metrics and operational scenarios [47]. The analysis reveals significant performance variations among different algorithmic approaches, providing empirical foundation for algorithm selection decisions in cross-border e-commerce fraud detection applications [48].

Overall Algorithm Performance Comparison

Classification Accuracy Metrics

Table 3 summarizes the primary classification performance metrics for each evaluated algorithm on the test dataset. The results represent average performance across five-fold cross-validation with 95% confidence intervals.

Table 3: Algorithm Performance Comparison on Test Dataset

Algorithm	Precision	Recall	F1-Score	AUC-ROC	False Positive Rate
Logistic Regression	0.742	0.685	0.712	0.876	0.031
Decision Tree	0.698	0.723	0.710	0.858	0.042
Random Forest	0.823	0.789	0.806	0.927	0.018
Gradient Boosting	0.847	0.812	0.829	0.941	0.015
XGBoost	0.861	0.826	0.843	0.948	0.013
MLP Neural Network	0.795	0.757	0.775	0.903	0.024
LSTM Network	0.778	0.741	0.759	0.891	0.027

The experimental results demonstrate clear performance hierarchies among algorithm classes [49]. Ensemble methods, particularly gradient boosting variants, achieve superior performance across all evaluation metrics compared to baseline algorithms and neural network approaches [50]. XGBoost attains the highest F1-score of 0.843 and AUC-ROC of 0.948, representing substantial improvements over logistic regression baseline [51]. The false positive rate of 0.013 for XGBoost translates to approximately 4,875 false alarms per 375,000 transactions in the test set, significantly lower than the 11,625 false alarms generated by logistic regression [52].

Random forest demonstrates strong performance with F1-score of 0.806, providing a favorable balance between accuracy and computational efficiency [53]. The ensemble nature of random forest contributes to robustness against overfitting and sensitivity to hyperparameter choices [54]. Neural network approaches, while theoretically capable of learning complex patterns, fail to match ensemble method performance in this fraud detection context [55]. The multilayer perceptron achieves respectable performance with F1-score of 0.775, while LSTM networks underperform expectations despite their sequential modeling capabilities [56]. The relatively modest dataset size and sparse fraud signals may limit neural network optimization through gradient-based training [57].

Precision-Recall Tradeoffs

The precision-recall tradeoff analysis examines algorithm behavior across different decision thresholds, revealing how threshold selection impacts operational

characteristics [58]. High precision settings minimize false positives at the cost of missing more fraud cases, suitable for scenarios where customer friction from false accusations carries high costs [59]. High recall settings maximize fraud detection at the expense of increased false positive rates, appropriate when preventing fraud takes precedence over operational efficiency [60]. Understanding these tradeoffs enables practitioners to calibrate fraud detection systems according to business priorities and risk tolerance [61].

Gradient boosting and XGBoost maintain superior precision-recall balance across the full threshold range [62]. At fixed recall of 0.80, XGBoost achieves precision of 0.854, substantially exceeding the 0.693 precision of logistic regression at equivalent recall [63]. This advantage translates to fewer false positives requiring manual review, reducing operational costs while maintaining detection coverage [64]. The area under the precision-recall curve, particularly informative for imbalanced classification problems, confirms ensemble method superiority with XGBoost scoring 0.837 compared to 0.694 for logistic regression [65]. Neural network methods exhibit more volatile precision-recall curves, suggesting sensitivity to threshold selection and potential calibration issues [66].

Feature Importance Analysis

Global Feature Importance

Feature importance analysis identifies which attributes contribute most significantly to fraud detection across all algorithms [67]. Table 4 presents the top 15 features ranked by average importance across ensemble methods, normalized to sum to 1.0.

Table 4: Top 15 Features by Importance Score

Rank	Feature Name	Importance Score	Category
1	Transaction Velocity (24h)	0.142	Temporal
2	Amount Deviation from History	0.118	Behavioral
3	IP-Billing Country Mismatch	0.095	Geographical
4	Account Age	0.087	User Attribute
5	Previous Decline Count	0.076	Transaction
6	Login Frequency Anomaly	0.068	Behavioral
7	Device Change Indicator	0.063	Device
8	Shipping-Billing Distance	0.059	Geographical
9	Time Since Last Transaction	0.054	Temporal
10	Merchant Risk Score	0.051	Transaction
11	Transaction Amount Percentile	0.048	Behavioral
12	Currency Mismatch	0.045	Transaction
13	Browser Fingerprint Change	0.042	Device
14	Hour of Day Risk Score	0.039	Temporal
15	Multiple Shipping Addresses	0.037	Behavioral

Transaction velocity emerges as the most discriminative feature, capturing rapid transaction sequences characteristic of card testing and account takeover fraud [68]. Amount deviation from historical patterns ranks second, identifying purchases significantly diverging from established spending behavior [69]. Geographical inconsistencies, particularly IP-billing country mismatches, provide strong fraud signals [70]. Account age contributes substantially to risk assessment, as newly created accounts exhibit higher fraud rates [71]. Previous decline history proves highly informative, as repeated payment failures often precede successful fraud attempts [72].

The feature importance distribution reveals dominance of behavioral and temporal features, collectively accounting for approximately 45% of total importance [73]. Geographical features contribute roughly 25%, while transaction attributes account for 20% [74]. Device-related features, despite comprising only 10% of total importance, provide valuable complementary information for fraud detection [75]. The relatively modest importance of individual device features suggests that device-based signals work synergistically with other feature categories rather than serving as standalone fraud indicators [76].

Feature Category Performance

Ablation studies assess the incremental contribution of different feature categories by systematically removing feature groups and measuring performance degradation [77]. The analysis reveals that removing temporal features reduces XGBoost F1-score from 0.843 to 0.761, representing the largest performance drop among all feature categories [78]. Removing behavioral features decreases F1-score to 0.778, confirming their substantial contribution [79]. Geographical feature removal reduces F1-score to 0.802, indicating moderate importance [80]. Transaction attribute removal yields F1-score of 0.815, while device feature removal results in minimal degradation to 0.837 [81]. These findings guide feature engineering priorities, suggesting that investments in refining temporal and behavioral features yield greater returns than equivalent efforts on device features [82].

Computational Efficiency Analysis

Practical fraud detection systems must balance prediction accuracy with computational resource requirements [83]. Real-time fraud prevention demands low-latency predictions compatible with typical payment processing timeframes [84]. Training efficiency affects how frequently models can be updated to adapt to evolving fraud patterns [85]. Table 5 quantifies computational characteristics for each evaluated algorithm.

Table 5: Computational Efficiency Metrics

Algorithm	Training Time (minutes)	Prediction Time (ms)	Memory Usage (MB)	Model Size (MB)
Logistic Regression	12.3	0.8	156	2.4
Decision Tree	8.7	1.2	189	5.8
Random Forest	47.5	4.3	1,247	142.6
Gradient Boosting	63.2	3.7	1,089	98.3
XGBoost	41.8	2.9	923	87.5
MLP Neural Network	92.4	5.1	2,134	215.7
LSTM Network	186.7	8.6	3,672	387.2

Logistic regression exhibits the fastest training and prediction times with minimal memory requirements, making it suitable for resource-constrained environments despite inferior accuracy [86]. XGBoost achieves favorable accuracy-efficiency balance, with training time of 41.8 minutes and prediction latency of 2.9 milliseconds per transaction [87]. The prediction latency remains well within acceptable bounds for real-time fraud detection, as typical payment processing allows several hundred milliseconds for fraud assessment [88]. Random forest requires more training time but offers competitive prediction speed [89]. Neural network approaches impose substantial computational burdens, with LSTM training time

exceeding three hours and prediction latency of 8.6 milliseconds, limiting their practical applicability [90].

Memory usage considerations affect deployment architecture decisions [91]. Ensemble methods require moderate memory allocation ranging from 900MB to 1,250MB during training [92]. Model sizes impact storage requirements and deployment flexibility, with XGBoost models consuming 87.5MB compared to 2.4MB for logistic regression [93]. The computational analysis suggests that XGBoost represents the optimal choice for most production fraud detection scenarios, offering superior accuracy without prohibitive resource requirements [94]. Organizations with stringent latency or resource constraints might accept lower accuracy from logistic regression or decision trees in exchange for operational simplicity [95].

Temporal Stability and Concept Drift

Fraud detection models must maintain performance over time despite evolving fraud patterns and legitimate customer behavior changes [96]. Temporal validation assesses algorithm stability by training on historical data and evaluating on subsequent time periods [97]. The analysis partitions the dataset into six two-month segments, training on the first four segments and testing on the final two segments representing future unseen data [98].

XGBoost maintains relatively stable performance with F1-score declining from 0.843 on concurrent test data to 0.814 on future data, representing a 3.4% degradation [99]. Gradient boosting exhibits similar stability with 3.8% performance decline [100]. Random forest demonstrates greater sensitivity to concept drift with 5.2% F1-score reduction [101]. Logistic regression suffers the most severe degradation at 8.7%, confirming its limited capacity to generalize beyond training distribution [102]. Neural network approaches show intermediate stability with 6-7% performance declines [103]. The temporal stability analysis indicates that ensemble methods maintain superior performance even under realistic deployment conditions involving concept drift [104].

Performance Under Class Imbalance

The severe class imbalance in fraud detection datasets poses algorithmic challenges [105]. Additional experiments evaluate algorithm performance under varying fraud rates by subsampling legitimate transactions to create synthetic datasets with fraud rates ranging from 1% to 10% [106]. XGBoost and gradient boosting maintain strong performance across all imbalance ratios, with F1-scores remaining above 0.80 even at 1% fraud rate [107]. Random forest exhibits moderate performance degradation at extreme imbalance ratios below 2% [108]. Neural network methods prove most sensitive to class imbalance, with F1-scores dropping below 0.70 at 1% fraud rate [109]. The robustness of ensemble methods to class imbalance stems from their hierarchical learning structure and built-in mechanisms for handling minority class samples [110].

Discussion and Practical Implications

The experimental findings provide actionable insights for developing and deploying fraud detection systems in cross-border e-commerce environments [111]. This section synthesizes key observations, discusses their practical implications, and offers recommendations for algorithm selection and system design [112].

Algorithm Selection Guidelines

The comprehensive performance evaluation reveals that XGBoost represents the optimal choice for most cross-border e-commerce fraud detection applications [113]. XGBoost achieves superior accuracy metrics including F1-score of 0.843 and AUC-ROC of 0.948, while maintaining acceptable computational efficiency with prediction latency of 2.9 milliseconds per transaction [114]. The false positive rate of 0.013 translates to manageable operational review workload, minimizing negative impacts on legitimate customer experiences [115].

Feature Engineering Recommendations

The feature importance analysis provides clear guidance for prioritizing feature engineering efforts. Temporal features, particularly transaction velocity metrics, emerge as the most valuable category for fraud detection. Organizations should invest in developing comprehensive temporal feature sets capturing transaction patterns across multiple time scales from minutes to months. Behavioral features derived from historical user activity constitute the second most important category, justifying expenditure on data infrastructure enabling efficient computation of aggregate statistics.

Conclusion

This research conducted systematic performance evaluation of machine learning algorithms for cross-border e-commerce fraud detection, providing empirical foundation for algorithm selection and system design decisions. The comprehensive experimental analysis encompassed multiple algorithm classes including traditional methods, ensemble techniques, and neural network approaches, evaluated across diverse performance metrics reflecting operational requirements. The findings reveal that ensemble methods, particularly XGBoost, achieve superior

accuracy while maintaining computational efficiency suitable for real-time fraud prevention systems.

The feature importance analysis identified temporal velocity metrics and behavioral deviation patterns as the most discriminative fraud indicators, guiding feature engineering priorities for organizations developing fraud detection capabilities. Geographical inconsistencies and account attributes provide additional valuable signals. The ablation studies quantified incremental contributions of different feature categories, enabling evidence-based resource allocation for feature development efforts. The research demonstrated that comprehensive feature coverage across multiple information sources outperforms sophisticated algorithms applied to limited feature sets.

Computational efficiency analysis confirmed that ensemble methods offer favorable accuracy-efficiency tradeoffs compared to neural network alternatives. XGBoost prediction latency of 2.9 milliseconds per transaction satisfies real-time processing requirements while delivering F1-score of 0.843 and false positive rate of 0.013. These performance characteristics enable effective fraud prevention without compromising legitimate customer experiences through excessive false alarms. The temporal stability evaluation verified that ensemble methods maintain performance under concept drift conditions typical of production environments.

The practical implications extend beyond algorithm selection to encompass holistic fraud detection system design. Organizations must consider feature engineering capabilities, computational infrastructure, interpretability requirements, regulatory compliance obligations, and operational constraints when formulating fraud detection strategies. The research provides actionable recommendations accommodating diverse organizational contexts and priorities. Future work should investigate adaptive learning mechanisms enabling continuous model improvement as fraud patterns evolve, explore privacy-preserving techniques reconciling detection effectiveness with data protection requirements, and extend evaluation to additional fraud types and e-commerce scenarios.

Acknowledgments

The authors express gratitude to the participating e-commerce platforms for providing anonymized transaction datasets essential for this research. We acknowledge the computational resources provided by the university research computing center that enabled extensive experimental evaluation. Special thanks to industry practitioners who offered valuable insights regarding practical fraud detection challenges and operational requirements. This work was supported by research grants focused on enhancing security in digital commerce ecosystems.

References

[1]. Guan, H., & Zhu, L. (2023). Dynamic Risk Assessment and Intelligent Decision Support System for Cross-border Payments Based on Deep Reinforcement Learning. Journal of Advanced Computing Systems, 3(9), 80-92.

[2]. Li, X., & Jia, R. (2024). Energy-aware scheduling algorithm optimization for AI workloads in data centers based on renewable energy supply prediction. Journal of Computing Innovations and Applications, 2(2), 56-65.

[3]. Yu, L., & Li, X. (2025). Dynamic optimization method for differential privacy parameters based on data sensitivity in federated learning. Journal of Advanced Computing Systems, 5(6), 1-13.

[4]. Weng, H., & Li, X. (2024). Renewable-Aware Cooperative Scheduling for Distributed AI Training Across Geo-Distributed Data Centers. Artificial Intelligence and Machine Learning Review, 5(2), 91-100.

[5]. Ye, H. (2024). Comparative Analysis of Deep Learning Algorithms for Disease-Related Protein Function Prediction: Performance Optimization and Computational Efficiency Evaluation. Artificial Intelligence and Machine Learning Review, 5(3), 80-97.

[6]. Ye, H. (2024). Cloud-based Data Mining for Cancer Drug Synergy Analysis: Applications in Non-small Cell Lung Cancer Treatment. Journal of Advanced Computing Systems, 4(4), 26-35.

[7]. Wang, Y., & Wang, X. (2023). FedPrivRec: A Privacy-Preserving Federated Learning Framework for Real-Time E-Commerce Recommendation Systems. Journal of Advanced Computing Systems, 3(5), 63-77.

[8]. Wang, Y. (2024). Comparative Analysis of AI-Driven Risk Prediction Methods in Retail Supply Chain Disruption Management: A Multi-Enterprise Study. Journal of Advanced Computing Systems, 4(4), 36-48.

[9]. Lu, X. (2025). DeepAd-OCR: An AI-Powered Framework for Automated Recognition and Enhancement of Conversion Elements in Digital Advertisements. Journal of Sustainability, Policy, and Practice, 1(4), 32-49.

[10]. Lu, X. (2024). Leveraging Generative AI for Cost-Effective Advertising Creative Automation: A

Practical Framework for Small and Medium Enterprises. Artificial Intelligence and Machine Learning Review, 5(2), 64-76.

[11]. Ge, L. (2023). Predictive Visual Analytics for Financial Anomaly Detection: A Big Data Framework for Proactive Decision Support in Volatile Markets. Artificial Intelligence and Machine Learning Review, 4(4), 42-56.

[12]. Pan, Z. (2025). A Reinforcement Learning Approach for Adaptive Budget Allocation in Pharmaceutical Digital Marketing: Maximizing ROI Across Patient Journey Touchpoints. Journal of Sustainability, Policy, and Practice, 1(4), 1-15.

[13]. Pan, Z. (2023). Machine Learning for Real-time Optimization of Bioprocessing Parameters: Applications and Improvements. Artificial Intelligence and Machine Learning Review, 4(3), 30-42.

[14]. Wu, C., & Pan, Z. (2024). An Integrated Graph Neural Network and Reinforcement Learning Framework for Intelligent Drug Discovery. Journal of Advanced Computing Systems, 4(6), 19-29.

[15]. Zhang, J. (2025). SecureCodeBERT: An Ai-Powered Model for Identifying and Categorizing High-Risk Security Vulnerabilities in Php-Based Critical Infrastructure Applications. Journal of Sustainability, Policy, and Practice, 1(4), 80-94.

[16]. Zhang, J. (2024). Evaluating Machine Learning Approaches for Sensitive Data Identification: A Comparative Study of NLP and Rule-Based Methods. Journal of Advanced Computing Systems, 4(7), 26-38.

[17]. Huang, Y. (2024). Fairness-Aware Credit Risk Assessment Using Alternative Data: An Explainable AI Approach for Bias Detection and Mitigation. Artificial Intelligence and Machine Learning Review, 5(1), 27-39.

[18]. Huang, Y. (2024). Graph-Based Feature Learning for Anti-Money Laundering in Cross-Border Transaction Networks. Journal of Advanced Computing Systems, 4(7), 39-49.

[19]. Lei, Y. (2025). RLHF-Powered Multilingual Audio Understanding: A Cross-Cultural Emotion Analysis Framework for International Communication. Journal of Sustainability, Policy, and Practice, 1(4), 66-79.

[20]. Cheng, Z. (2024). Attention-Enhanced Multi-Scale Feature Optimization for Silent Myocardial Infarction and Early Atrial Fibrillation Detection in ECG Signals. Artificial Intelligence and Machine Learning Review, 5(3), 67-79.

[21]. Cai, Y. (2025). Federated Learning-Based Framework for Privacy-Protected Cross-Border Financial Risk Evaluation: Analyzing US-Asia Investment Flows. Journal of Sustainability, Policy, and Practice, 1(4), 50-65.

[22]. Cai, Y. (2023). Multi-Horizon Financial Crisis Detection Through Adaptive Data Fusion. Artificial Intelligence and Machine Learning Review, 4(1), 16-30.

[23]. Cai, Y. (2024). Comparative Evaluation of Feature Extraction Techniques in Margin Call Cascade Detection: Balancing Accuracy and False Alarm Rates. Journal of Advanced Computing Systems, 4(7), 1-12.

[24]. Long, X. (2024). Optimizing Deep Learning Algorithms for Enhanced Detection Accuracy in Distributed Network Attack Scenarios. Artificial Intelligence and Machine Learning Review, 5(1), 79-92.

[25]. Liu, Y. (2025). Research on AI Driven Cross Departmental Business Intelligence Visualization Framework for Decision Support. Journal of Sustainability, Policy, and Practice, 1(2), 69-85.

[26]. Wang, J. (2024). Multimodal Deep Learning Approach for Early Warning of Supply Chain Disruptions Using NLP and Anomaly Detection. Artificial Intelligence and Machine Learning Review, 5(3), 98-110.

[27]. Wang, Z. (2024). Adaptive Ensemble Learning Framework with SHAP-Based Feature Optimization for Financial Anomaly Detection. Artificial Intelligence and Machine Learning Review, 5(1), 51-66.

[28]. Wang, Z. (2024). Enhancing Financial Named Entity Recognition through Adaptive Few-Shot Learning: A Comparative Study of Pre-trained Language Models. Journal of Advanced Computing Systems, 4(7), 13-25.

[29]. Dong, Z. (2024). Adaptive UV-C LED Dosage Prediction and Optimization Using Neural Networks Under Variable Environmental Conditions in Healthcare Settings. Journal of Advanced Computing Systems, 4(3), 47-56.

[30]. Dong, Z. (2024). AI-Driven Reliability Algorithms for Medical LED Devices: A Research Roadmap. Artificial Intelligence and Machine Learning Review, 5(2), 54-63.

[31]. Li, J., Ren, W., & Wu, X. (2023). Early Malware Detection through Temporal Analysis of System Behaviors. Journal of Global Engineering Review, 1(1), 1-11.

[32]. Li, J., Ren, W., & Wu, X. (2024). Semi-Supervised Learning Approach for Automated Sensitive Data Classification in Unstructured Text Documents. Journal of Global Engineering Review, 2(2), 1-17.

[33]. Wu, X., Li, J., & Ren, W. (2024). Risk Assessment Framework for Data Leakage Prevention Using Machine Learning Techniques. Artificial Intelligence and Machine Learning Review, 5(3), 55-66.

[34]. Ren, W., Li, J., & Wu, X. (2024). Privacy-Preserving Data Analysis Using Federated Learning: A Practical Implementation Study. Artificial Intelligence and Machine Learning Review, 5(1), 40-50.

[35]. Weng, H., Zhang, S., & Min, S. (2024). Multi-Constraint Optimization for Real-Time Bidding: A Reinforcement Learning Approach. Artificial Intelligence and Machine Learning Review, 5(1), 93-104.

[36]. Zhang, S., Wang, Y., & Weng, H. (2024). Industrial IoT Anomaly Detection Using Improved Autoencoder Architecture. Artificial Intelligence and Machine Learning Review, 5(1), 67-78.

[37]. Weng, H., Wang, H., & Wei, C. (2024). Adaptive Bidding Strategies for Hybrid Auction Mechanisms in Programmatic Advertising. Journal of Advanced Computing Systems, 4(4), 13-25.

[38]. Weng, H., & Li, X. (2024). Renewable-Aware Cooperative Scheduling for Distributed AI Training Across Geo-Distributed Data Centers. Artificial Intelligence and Machine Learning Review, 5(2), 91-100.

[39]. Kang, A., Xin, J., & Ma, X. (2024). Anomalous cross-border capital flow patterns and their implications for national economic security: An empirical analysis. Journal of Advanced Computing Systems, 4(5), 42-54.

[40]. Kang, A., Li, Z., & Meng, S. (2023). AI-Enhanced Risk Identification and Intelligence Sharing Framework for Anti-Money Laundering in Cross-Border Income Swap Transactions. Journal of Advanced Computing Systems, 3(5), 34-47.

[41]. Kang, A., Min, S., & Yuan, D. (2024). Comparative Analysis of Foreign Exchange Market Shock Transmission and Recovery Resilience Among Major Economies Under Geopolitical Conflicts: Evidence from the Russia-Ukraine Crisis. Journal of Computing Innovations and Applications, 2(1), 46-61.

[42]. Dong, B., Zhang, D., & Xin, J. (2024). Deep reinforcement learning for optimizing order book imbalance-based high-frequency trading strategies. Journal of Computing Innovations and Applications, 2(2), 33-43.

[43]. Xiong, K., Wu, Z., & Jia, X. (2025). Deepcontainer: a deep learning-based framework for real-time anomaly detection in cloud-native container environments. Journal of Advanced Computing Systems, 5(1), 1-17.

[44]. Wu, Z., Wang, S., Ni, C., & Wu, J. (2024). Adaptive traffic signal timing optimization using deep reinforcement learning in urban networks. Artificial Intelligence and Machine Learning Review, 5(4), 55-68.

[45]. Wu, Z., Feng, E., & Zhang, Z. (2024). Temporal-Contextual Behavioral Analytics for Proactive Cloud Security Threat Detection. Academia Nexus Journal, 3(2).

[46]. Wu, Z., Feng, Z., & Dong, B. (2024). Optimal feature selection for market risk assessment: A dimensional reduction approach in quantitative finance. Journal of Computing Innovations and Applications, 2(1), 20-31.

[47]. Zhang, Z., & Wu, Z. (2023). Context-aware feature selection for user behavior analytics in zero-trust environments. Journal of Advanced Computing Systems, 3(5), 21-33.

[48]. Li, J., Ren, W., & Wu, X. (2024). Semi-Supervised Learning Approach for Automated Sensitive Data Classification in Unstructured Text Documents. Journal of Global Engineering Review, 2(2), 1-17.

[49]. Temporal Feature Analysis of Transaction Sequences for Payment Fraud Identification in Small and Medium-Sized Enterprises

[50]. Li, J., Ren, W., & Wu, X. (2025). Temporal Feature Analysis of Transaction Sequences for Payment Fraud Identification in Small and Medium-Sized Enterprises. Journal of Global Engineering Review, 3(1), 1-18.

[51]. Ren, W., Wu, X., & Li, J. (2025). AI-Driven Network Threat Behavior Pattern Recognition and Classification: An Ensemble Learning Approach with Temporal Analysis. Journal of Advanced Computing Systems, 5(9), 1-13.

[52]. Wu, X., Li, J., & Ren, W. (2024). Risk Assessment Framework for Data Leakage Prevention Using Machine Learning Techniques. Artificial Intelligence and Machine Learning Review, 5(3), 55-66.

[53]. Ren, W., Li, J., & Wu, X. (2024). Privacy-Preserving Data Analysis Using Federated Learning: A Practical Implementation Study. Artificial Intelligence and Machine Learning Review, 5(1), 40-50.

[54]. [114]Tu, W., Wan, G., Shang, Z., & Du, B. (2025). Efficient relational context perception for knowledge graph completion. Applied Intelligence, 55(15), 1005.

[55]. Weng, H., Zhang, S., & Min, S. (2024). Multi-Constraint Optimization for Real-Time Bidding: A Reinforcement Learning Approach. Artificial Intelligence and Machine Learning Review, 5(1), 93-104.

[56]. Zhang, S., Wang, Y., & Weng, H. (2024). Industrial IoT Anomaly Detection Using Improved Autoencoder Architecture. Artificial Intelligence and Machine Learning Review, 5(1), 67-78.

[57]. Adaptive Bidding Strategies for Hybrid Auction Mechanisms in Programmatic Advertising

[58]. Wu, C., Guan, H., & Weng, H. (2024). Forecasting Hospital Resource Demand Using Gradient Boosting: An Operational Analytics Approach for Bed Allocation and Patient Flow Management. Journal of Computing Innovations and Applications, 2(1), 74-85.

[59]. Weng, H., & Li, X. (2024). Renewable-Aware Cooperative Scheduling for Distributed AI Training Across Geo-Distributed Data Centers. Artificial Intelligence and Machine Learning Review, 5(2), 91-100.

[60]. Weng, H., & Lei, Y. (2024). Cross-Modal Artifact Mining for Generalizable Deepfake Detection in the Wild. Journal of Computing Innovations and Applications, 2(2), 78-87.

[61]. Shi, X., & Weng, H. (2024). Comparative Analysis of Unsupervised Learning Approaches for Anomalous Billing Pattern Detection in Healthcare Payment Integrity. Journal of Computing Innovations and Applications, 2(1), 111-127.

[62]. Weng, H. (2025). Deep Embedding Clustering with Adaptive Feature Selection for Banking Customer Segmentation. Spectrum of Research, 5(2).

[63]. Li, J., Ren, W., & Wu, X. (2023). Early Malware Detection through Temporal Analysis of System Behaviors. Journal of Global Engineering Review, 1(1), 1-11.