DOI 10.34185/1562-9945-5-160-2025-16
UDC 004.8

K.Y. Ostrovska, V.O. Nosov

MACHINE LEARNING METHODS FOR ANTIFRAUD SYSTEMS

Annotation. Fraud in the financial sector, e-commerce, and online services is becoming increasingly frequent and sophisticated. Traditional rule-based systems, while still helpful in detecting known fraud patterns, struggle to keep up with new, evolving attack vectors, as static rules are quickly circumvented. In contrast, machine learning (ML) provides a dynamic and scalable approach that can process vast amounts of transactional and behavioral data to identify subtle anomalies and suspicious activity.

This paper provides a comprehensive overview of current ML techniques used in fraud detection, categorized into three main groups: classification models, anomaly detection methods, and deep learning architectures. It discusses real-world applications across various fraud scenarios, including credit card abuse, account takeovers, cybercrime, and scams in digital comerce.

Emphasis is placed on the strengths and limitations of each approach, with attention to real-world considerations like scalability, model transparency, and the challenge of class imbalance. The paper also reviews recent advances, including graph-based representations of financial interactions, IP-based behavioral profiling, and the emergence of hybrid systems that integrate multiple ML techniques – such as combining autoencoders with boosting algorithms for improved accuracy, especially when labeled data is scarce.

The findings aim to support the development of flexible, high-performance fraud detection solutions that leverage the most effective ML practices and capitalize on the synergy of hybrid model architectures.

Keywords: Fraud detection, machine learning, classification, anomaly detection, neural networks, hybrid approaches.

Problem statement. Fraudsters constantly evolve their methods, complicating the maintenance of traditional rule-based systems and reducing their effectiveness. ML models can adapt to new fraud patterns such as identity theft, account takeovers, and money laundering and scale efficiently with growing data volumes, making them tools for fraud detection.

Fraud takes on many guises, from credit card fraud to account takeovers and online scams. Each form requires a unique strategy for effective detection. Financial fraud often involves unauthorized transactions, identity theft, or money laundering, while cyber and e-commerce fraud may exploit system vulnerabilities and social engineering techniques. With the increasing prevalence of digital banking and authentication-based services, account-related fraud has become a particularly relevant area for specialized strategies.

ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)

Recent statistics from the National Bank of Ukraine [1] highlight the evolving nature of payment fraud. Although the number of fraudulent card transactions in Ukraine slightly declined by 1% in 2024 compared to the previous year, the total financial losses caused by such fraud surged by 37%, exceeding 1.1 billion UAH. Furthermore, more than 83% of these fraudulent operations were carried out remotely – often through online purchases, fake service payments, and manipulative schemes such as phishing or social engineering. This ongoing shift toward digital fraud emphasizes the growing importance of behavioral and contextual risk analysis, which cannot be addressed by static rule-based systems alone and requires the implementation of adaptive detection strategies.

Modern fraud detection systems must possess three key attributes: scalability, adaptability, and real-time data processing. ML algorithms enable systems to learn from historical and behavioral data, detect rare events, and continuously improve. One of the most powerful aspects of these algorithms is their ability to uncover hidden correlations, providing a deeper understanding of fraud patterns and tactics. Combined with its adaptability, it ensures that models can quickly respond to evolving fraud tactics.

According to BioCatch [2], over 83% of financial institutions worldwide already use ML to prevent fraud. A prominent example of such technology adoption in large-scale business is the Visa payment system, which is actively developing innovative AI-based solutions, including generative AI [3], a type of AI that can generate new data based on patterns in existing data to counter sophisticated fraud schemes in digital banking.

Therefore, reviewing effective methodologies is essential for understanding the potential directions for developing advanced fraud prevention mechanisms.

Classification algorithms. Most financial fraud detection tasks are formulated as binary classification problems (“fraud” vs “normal”). Among classical approaches, logistic regression and decision trees are widely used. Logistic regression is a simple and interpretable model where coefficients can be interpreted as feature weights. However, due to its linear nature, it cannot capture complex, non-linear interactions between features, which limits its effectiveness in detecting sophisticated fraud schemes, such as money laundering or insider trading. Decision trees, in turn, generate explicit “if-then” rules that are easily understandable by experts, but individual trees tend to overfit and generally underperform compared to ensemble methods [4]. In other words, while logistic regression and decision trees are commonly used as baseline models, their performance in complex scenarios is typically inferior to more modern approaches.

Ensemble classifiers are increasingly applied to improve fraud detection accuracy, particularly bagging and boosting methods. A typical example of bagging is the Random Forest (RF), which combines many decision trees. By averaging results across many models, RF achieves better generalization and robustness to noise and outliers in the data. Studies show that RF outperforms individual trees in financial fraud detection tasks [4]. Its drawback lies in relatively low interpretability: explaining results from an ensemble of hundreds of trees can be challenging. On the other hand, gradient boosting methods build a sequence of trees that progressively correct errors from previous iterations, allowing the modeling of highly complex dependencies. Modern boosting implementations, such as XGBoost (Extreme Gradient Boost-

ing), LightGBM (Light Gradient Boosting Machine), and CatBoost (Category Boost), have demonstrated high performance in fraud detection tasks and are well-suited for class imbalance [4].

Specifically, XGBoost has proven effective on highly imbalanced data, which is common in fraudulent transaction detection. Class imbalance refers to the situation where the number of instances of one class is significantly higher than that of the other, making it challenging for the model to learn from the minority class. XGBoost includes built-in mechanisms for class weight compensation [4][8] to address this issue.

In addition to class weighting, oversampling techniques have also been used to improve performance on imbalanced datasets. One of the most widely adopted methods is SMOTE (Synthetic Minority Oversampling Technique). SMOTE generates synthetic examples of the minority class by interpolating between existing instances, effectively increasing the diversity of fraudulent samples in the training data. This approach has shown strong results in fraud detection contexts. For instance, in a recent study [5], the combination of RF with SMOTE achieved up to 99.5% accuracy on a dataset with less than 0.2% fraud cases. Similarly, practical applications such as Kaggle models demonstrate that SMOTE significantly increases recall, capturing more fraudulent transactions while maintaining acceptable levels of precision [6]. However, it is essential to apply SMOTE only on the training set to avoid information leakage into the test set.

Other boosting frameworks have their own advantages: LightGBM typically trains faster due to its leafwise growth strategy, while CatBoost can automatically handle categorical features, reducing the need for manual preprocessing. Overall, ensemble methods are currently among the most effective techniques for fraud detection. Their typical limitations include high computational complexity and the black-box nature of model decisions, which can be a challenge in highly regulated financial sectors.

Anomaly detection algorithms. In scenarios lacking labeled data or needing to identify previously unknown fraud patterns, anomaly detection methods (unsupervised) are commonly used. These approaches treat fraud cases as statistically rare anomalies that deviate from “normal” transaction profiles. The goal is to model everyday transactions and detect those that do not conform to this pattern. Classic algorithms include Isolation Forest (IF), One-Class Support Vector Machines (SVM), Local Outlier Factor (LOF), and clustering-based methods such as k-means.

IF is a variant of RF specifically adapted to isolate anomalous points: randomly generated trees can quickly separate “unusual” records with outlier feature values. This method scales well to large datasets and has been successfully used for unsupervised fraud detection [8][9]. Researchers report that IF can detect anomalous payment transactions with high accuracy, as indicated by an AUC value of approximately 0.82 on simulated test sets [9].

One-Class SVM constructs a hyperplane that encompasses most normal data and classifies points outside as anomalies. It has been applied in fraud detection, although its performance is sensitive to kernel choice and parameter tuning, and it scales poorly with high-dimensional data. LOF evaluates the local density around each point – transactions with sig-

ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)

nificantly lower density compared to neighbors are flagged as potential anomalies. LOF can detect local outliers but has limited scalability due to the computational cost of distance calculations across large datasets.

Clustering methods like k-means, followed by identifying points far from cluster centroids, provide a simple alternative, although the effectiveness of this approach depends on the assumption that normal data form compact clusters, which is not always the case.

Autoencoders, a type of deep neural network trained to reconstruct input data, deserve special attention. When trained only on normal transactions, autoencoders reconstruct anomalous (fraudulent) examples poorly, resulting in a high reconstruction error. By comparing this error to a threshold, suspicious operations can be flagged. Recent studies report the successful use of autoencoders for fraud detection. For instance, work [8] proposes a neural network based on an attention-guided autoencoder combined with a Generative Adversarial Network (GAN), a type of neural network that learns to generate data similar to the training data. This combined model, trained solely on normal data, effectively separated fraudulent records as anomalies, even in highly imbalanced datasets [8].

Another recent study [7] demonstrates that combining an autoencoder with gradient boosting significantly improves detection performance: the autoencoder is used for dimensionality reduction and feature transformation, followed by a LightGBM classifier. The hybrid autoencoders + LightGBM model achieved a recall of ~94.8%, which measures the proportion of actual fraud cases that were correctly identified, significantly outperforming standalone models (with the best alternative recall of around 86%) on the same credit transaction dataset [7]. This approach merges the strengths of unsupervised anomaly detection (the ability to discover novel fraud patterns) with the precision of supervised classification. Similar hybrid architectures have proven effective in other research as well; for instance, XGBOD (Extreme Boosting Outlier Detection), a framework that uses multiple anomaly detectors as feature generators for a gradient boosting model, outperformed single-model baselines in fraud detection benchmarks [6].

In addition to the methods described above, modeling the temporal nature of transactions has proven to be an effective approach for spotting anomalies that unfold over time. Fraudulent activity often shows up as sudden, unexpected changes in how and when transactions occur – patterns that may be missed by models treating each transaction in isolation. To capture such dynamics, Recurrent Neural Networks (RNNs) and their more advanced variant, Long Short-Term Memory (LSTM) networks, have become increasingly popular. These architectures are designed to process sequences of data while preserving context from earlier steps, enabling them to recognize unusual transaction flows that differ from a user's typical usage patterns.

For example, the study [10] demonstrated that using multiple interleaved RNNs to model parallel usage streams (time, device, location) improved fraud detection accuracy while reducing the need for handcrafted features. LSTM networks, in particular, are capable of capturing long-term dependencies in user transaction activity, making them well-suited for identifying subtle shifts in patterns that static models may miss. When combined with attention

mechanisms, these models not only improve accuracy but also enhance interpretability by highlighting the most relevant portions of the transaction sequence.

Thus, anomaly detection techniques, especially when extended with sequence-based neural models, are beneficial for detecting underrepresented or emerging fraud scenarios, especially when labeled data is limited.

Deep learning and specialized methods. The growth of data volume and the increasing complexity of fraud schemes have led to the adoption of deep learning. Multilayer neural networks can automatically capture hidden relationships in transactional data. When large historical datasets are available, deep models (feedforward neural networks or recurrent neural networks) often outperform traditional methods in accuracy. In some experiments, deep convolutional neural networks (CNNs) improved fraud detection over tree-based ensembles when applied to large transaction volumes. However, neural networks require careful tuning and longer training times and are susceptible to overfitting on non-representative data.

A cutting-edge direction involves modeling inter-object relationships in financial data through Graph Neural Networks (GNNs). Fraud often involves complex network structures, including links between customers, merchants, devices, or IP addresses. GNNs enable learning directly on graph-structured data, leveraging the interconnectivity of nodes. A recent review [11] demonstrates that GNNs are particularly effective for detecting fraud in financial networks, as they capture hidden patterns of interaction (such as fraudulent transaction rings), significantly outperforming traditional models in accuracy. In real-world systems, GNNs are already used to analyze payment graphs and identify suspicious subgraphs – an approach that would be difficult for models lacking structural awareness.

Another industrial approach involves methods like IP Insights, which are designed to detect IP-based anomalies. For example, Amazon's IP Insights algorithm is trained unsupervised to learn behavior profiles from historical user-IP pairs and assess the "unlikeliness" of new login attempts [12]. This enables the detection of abnormal behavior – for example, when a user account suddenly has many new IP addresses or logins from unexpected regions, possibly indicating an account compromise.

Conclusions. This study provides a foundation for developing effective fraud detection systems utilizing modern ML techniques. It provides an overview of key approaches, from basic classification models and anomaly detection methods to more advanced deep learning architectures. It demonstrates how ML can enable adaptive, scalable, and data-driven solutions in the fight against fraud.

Ensemble classification methods, particularly gradient boosting, remain widely used due to their ability to model complex patterns and handle class imbalance. Anomaly detection techniques, such as autoencoders and IF, help uncover unusual or previously unknown fraud cases without requiring prior data labeling. Deep learning introduces new capabilities, enabling it to detect more complex patterns and account for temporal dependencies in transaction streams.

Specialized approaches such as GNNs and IP-based behavioral profiling open new prospects for detecting fraud embedded in complex relationships or device-level signals. Hybrid

ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)

approaches that combine different methods, for example, using autoencoders for feature extraction followed by gradient boosting, have demonstrated high effectiveness on real-world datasets with class imbalance.

Today, the process of fraud detection is increasingly focused on building flexible, intelligent systems that integrate multiple methods while addressing practical concerns such as transparency, fairness, and real-time responsiveness. To stay ahead of emerging threats, these systems must incorporate continuous learning, real-time data stream processing, and human involvement as part of a robust and adaptive defense strategy.

ЛІТЕРАТУРА

1. Національний Банк України. Кількість випадків шахрайства з картками знизилася, збитки за ними – зросли. 12.05.2025. URL: https://bank.gov.ua/ua/news/all/kilkist-vipadkiv-shahraystva-z-kartkami-znizilasya-zbitki-za-nimi--zrosli (дата звернення: 16.06.2025).
2. 2024 AI, Fraud, and Financial Crime Survey. URL: https://www.biocatch.com/ai-fraud-financial-crime-survey (дата звернення: 14.06.2025).
3. Visa Announces Generative AI-Powered Fraud Solution to Combat Account Attacks. 07.05.2024. URL: https://investor.visa.com/news/news-details/2024/Visa-Announces-Generative-AI-Powered-Fraud-Solution-to-Combat-Account-Attacks/default.aspx (дата звернення: 14.06.2025).
4. Big Data-Driven Distributed Machine Learning for Scalable Credit Card Fraud Detection Using PySpark, XGBoost, and CatBoost / T. Leonidas та ін. *Electronics 2025*. 2025. Т. 14. С. 1754. URL: https://doi.org/10.3390/electronics14091754.
5. Optimizing credit card fraud detection with random forests and SMOTE / P. Sundaravadivel та ін. *Scientific Reports*. 2025. Т. 15. С. 17851. URL: https://doi.org/10.1038/s41598-025-00873-y.
6. Applying SMOTE to Fraud Detection. *Kaggle*. URL: https://www.kaggle.com/code/wuttipats/applying-smote-to-fraud-detection (дата звернення: 14.06.2025).
7. An AutoEncoder enhanced light gradient boosting machine method for credit card fraud detection / L. Ding та ін. *PeerJ Computer Science*. 2024. Т. 10. С. e2323. URL: https://doi.org/10.7717/peerj-cs.2323 (дата звернення: 15.06.2025).
8. Credit Card Fraud Detection Based on Unsupervised Attentional Anomaly Detection Network / S. Jiang та ін. *Systems*. 2023. Т. 11, № 6. С. 305. URL: https://doi.org/10.3390/systems11060305 (дата звернення: 15.06.2025).
9. Buschjäger S., Honysz P.-J., Morik K. Randomized outlier detection with trees. *International Journal of Data Science and Analytics*. 2020. URL: https://doi.org/10.1007/s41060-020-00238-w (дата звернення: 15.06.2025).
10. Interleaved Sequence RNNs for Fraud Detection / B. Branco та ін. *KDD '20: The 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, м. Virtual Event CA USA. New York, NY, USA, 2020. URL: https://doi.org/10.1145/3394486.3403361 (дата звернення: 16.06.2025).

11. Graph neural networks for financial fraud detection: a review / D. Cheng та ін. *Frontiers of Computer Science*. 2025. Т. 19, № 9. URL: https://doi.org/10.1007/s11704-024-40474-y (дата звернення: 18.06.2025).

12. How IP Insights Works - Amazon SageMaker AI. URL: https://docs.aws.amazon.com/sagemaker/latest/dg/ip-insights-howitworks.html (дата звернення: 16.06.2025).

REFERENCES

1. National Bank of Ukraine. (2025). The number of card fraud cases has decreased, but the losses from them have increased. https://bank.gov.ua/ua/news/all/kilkist-vipadkiv-shahraystva-z-kartkami-znizilasya-zbitki-za-nimi--zrosli.

2. *2024 AI Fraud Financial Crime Survey*. (2025). https://www.biocatch.com/ai-fraud-financial-crime-survey.

3. Visa Announces Generative AI-Powered Fraud Solution to Combat Account Attacks. (2024). https://investor.visa.com/news/news-details/2024/Visa-Announces-Generative-AI-Powered-Fraud-Solution-to-Combat-Account-Attacks/default.aspx.

4. Theodorakopoulos, L., Theodoropoulou, A., Tsimakis, A., & Halkiopoulos, C. (2025). Big Data-Driven Distributed Machine Learning for Scalable Credit Card Fraud Detection Using PySpark, XGBoost, and CatBoost. *Electronics*, *14*(9), 1754. https://doi.org/10.3390/electronics14091754.

5. Sundaravadivel, P., Isaac, R., Elangovan, D., KrishnaRaj, D., Rahul, V., & Raja, R. (2025). Optimizing credit card fraud detection with random forests and SMOTE. *Scientific Reports*, *15*, 17851. https://doi.org/10.1038/s41598-025-00873-y.

6. *Applying SMOTE to Fraud Detection*. (2023). Kaggle. https://www.kaggle.com/code/wuttipats/applying-smote-to-fraud-detection.

7. Ding, L., Liu, L., Wang, Y., Shi, P., & Yu, J. (2024). An AutoEncoder enhanced light gradient boosting machine method for credit card fraud detection. *PeerJ Computer Science*, *10*, Article e2323. https://doi.org/10.7717/peerj-cs.2323.

8. Jiang, S., Dong, R., Wang, J., & Xia, M. (2023). Credit Card Fraud Detection Based on Unsupervised Attentional Anomaly Detection Network. *Systems*, *11*(6), 305. https://doi.org/10.3390/systems11060305.

9. Buschjäger, S., Honysz, P.-J., & Morik, K. (2020). Randomized outlier detection with trees. *International Journal of Data Science and Analytics*. https://doi.org/10.1007/s41060-020-00238-w.

10. Branco, B., Abreu, P., Gomes, A. S., Almeida, M. S. C., Ascensão, J. T., & Bizarro, P. (2020). Interleaved Sequence RNNs for Fraud Detection. У *KDD '20: The 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. ACM. https://doi.org/10.1145/3394486.3403361.

11. Cheng, D., Zou, Y., Xiang, S., & Jiang, C. (2025). Graph neural networks for financial fraud detection: a review. *Frontiers of Computer Science*, *19*(9). https://doi.org/10.1007/s11704-024-40474-y.

ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)

12. *How IP Insights Works - Amazon SageMaker AI*. (2025). https://docs.aws.amazon.com/sagemaker/latest/dg/ip-insights-howitworks.html.

Received 18.06.2025.
Accepted 24.06.2025.

Методи машинного навчання для антифрод-систем

Анотація. Шахрайство у фінансовому секторі, сфері електронної комерції та онлайн-сервісах стає дедалі частішим і витонченішим. Традиційні системи на основі правил, хоча й залишаються корисними для виявлення відомих шаблонів шахрайства, не встигають за новими, динамічними схемами атак, оскільки статичні правила швидко обходяться зловмисниками. Натомість машинне навчання (МН) пропонує динамічний і масштабований підхід, здатний обробляти великі обсяги транзакційних і поведінкових даних для виявлення тонких аномалій та підозрілої активності.

У статті представлено ґрунтовний огляд сучасних методів МН, що застосовуються для виявлення шахрайства. Вони згруповані у три основні категорії: моделі класифікації, методи виявлення аномалій та глибинні архітектури. Розглянуто приклади практичного використання в різноманітних сценаріях шахрайства, зокрема зловживання кредитними картками, перехоплення облікових записів, кіберзлочини та шахрайські дії в цифровій торгівлі.

Особливу увагу приділено перевагам і обмеженням кожного підходу з урахуванням таких практичних аспектів, як масштабованість, прозорість моделей і проблема дисбалансу класів. Також проаналізовано останні досягнення у цій сфері, зокрема використання графових представлень фінансових взаємодій, поведінкове профілювання на основі IP-адрес, а також поява гібридних систем, які поєднують декілька методів МН, наприклад, автоенкодери з бустинговими алгоритмами, які використано для підвищення точності, особливо у випадках нестачі розмічених даних.

Результати дослідження спрямовані на підтримку розробки гнучких, високоефективних систем виявлення шахрайства, які використовують найкращі практики МН та поєднують переваги гібридної архітектури моделей.

Ключові слова: виявлення шахрайства, машинне навчання, класифікація, виявлення аномалій, нейронні мережі, гібридні підходи.

Островська Катерина Юріївна – к.т.н., доцент кафедри інформаційних технологій і систем Українського державного університету науки і технологій.
Носов Валерій Олександрович – аспірант кафедри інформаційних технологій і систем Українського державного університету науки і технологій.

Ostrovska Kateryna Yuriivna – Ph.D., Associate Professor of the Department of Information Technology and Systems of Ukrainian State University of Science and Technology.
Nosov Valerii Oleksandrovych – postgraduate student of the Department of Information Technology and Systems of Ukrainian State University of Science and Technology.