UDC 004.89:336.74

DOI: 10.15587/1729-4061.2025.335630

ENHANCED IDENTIFICATION OF ILLICIT BITCOIN TRANSACTIONS THROUGH GENETIC ALGORITHM-BASED FEATURE SELECTION

Medet Shaizat*

Shynar Mussiraliyeva
Corresponding author
PhD*
E-mail: mussiraliyevash@gmail.com

*Department of Cybersecurity and Cryptology
Al-Farabi Kazakh National University
Al-Farabi ave., 71, Almaty,
Republic of Kazakhstan, 050040

The object of this study is the process of classifying illicit Bitcoin transactions in blockchain datasets. The problem addressed in this work is the difficulty of detecting suspicious activity in cryptocurrency networks due to the high dimensionality of transaction data and the lack of semantic labels, which limits the effectiveness of conventional manual feature engineering. The proposed method combines domain-specific indicators of illicit behavior with a Genetic Algorithm-driven selection mechanism that dynamically evolves informative feature subsets. The developed framework was implemented and evaluated on the Elliptic and Elliptic++ datasets using random forest. The results obtained demonstrate that the GA-based method significantly increases model performance: the best-performing configuration achieved an F1-score of 84.3%, a precision of 99.4%, and a recall of 73.1%. Compared to baseline approaches on the same dataset, this method provides relative improvements of 0.9% in F1-score, 0.3% in precision, and 1.2% in recall. The effectiveness of the proposed solution is explained by its ability to detect hidden patterns in transactional data with many potential attributes without resorting to manual heuristics, as well as an optimized setting of Genetic Algorithm parameters. A distinctive feature of this method is the combination of heuristic search with domain-informed feature categories, which improves classification accuracy and reduces model complexity. The obtained results can be applied in practical scenarios such forensic analysis of cryptocurrency transactions. However, successful implementation requires access to historical transaction records and sufficient computing resources to process large, feature-rich datasets

Keywords: bitcoin, illicit transactions, machine learning, genetic algorithms, cryptocurrency forensics

Received 06.05.2025
Received in revised form 28.06.2025
Accepted date 17.07.2025
Published date 29.08.2025

How to Cite: *Shaizat, M., Mussiraliyeva, S. (2025). Enhanced identification of illicit bitcoin transactions through genetic algorithm-based feature selection. Eastern-European Journal of Enterprise Technologies, 4 (9 (136)), 34–42. https://doi.org/10.15587/1729-4061.2025.335630*

1. Introduction

Cryptocurrencies are a new class of digital assets that operate independently of centralized banking systems and government regulation [1]. Among them, Bitcoin, introduced in 2009, has become the most prominent example, combining cryptographic protocols with a decentralized blockchain infrastructure to enable peer-to-peer transactions with low associated costs [2, 3]. While these technological innovations have demonstrated significant benefits in terms of transparency, efficiency, and financial inclusion, they have also opened new opportunities for illicit financial activity [4].

One of the major concerns surrounding Bitcoin is its potential use in illicit transactions due to the pseudonymous nature of blockchain records. While all transactions are publicly recorded on the ledger, the lack of direct identifying information poses significant challenges for law enforcement and compliance agencies [5]. Recent reports have shown that terrorist organizations, including internationally sanctioned groups, are increasingly using cryptocurrencies to finance activities such as fundraising, sanctions evasion, cybercrime, and internal value transfers [5, 6]. According to the Chainalysis (2025) report, in 2024, illicit cryptocurrency addresses received approximately 40.9 billion USD, with estimates suggesting the total could exceed 51 billion USD once all transactions are accounted for. This marks a significant increase from the previous year, indicating a growing trend in crypto-related crime [7].

The detection and prevention of illicit transactions in Bitcoin networks is complicated by the unique properties of blockchain data. Unlike traditional banking systems, all Bitcoin transactions are publicly recorded in an immutable ledger accessible to anyone [8]. Despite this transparency, Bitcoin remains pseudonymous: wallet addresses are not directly tied to personal identities, making it challenging to trace funds to individuals or organizations. This paradox-full visibility of transaction flows yet limited visibility of actors-poses a significant obstacle for regulatory and compliance efforts.

In recent years, researchers have proposed a variety of analytical approaches to detect suspicious activity in cryptocurrency networks. These approaches include network topology analysis, heuristic address clustering, and increasingly, the application of machine learning methods capable of analyzing transaction patterns at scale. Machine learning classifiers, whether supervised, unsupervised, or semi-supervised have demonstrated promise in identifying anomalous behavior by exploiting features such as transaction frequency, value distribution, and address connectivity. Nonetheless, a persistent problem in this domain is the high dimensionality of transaction data combined with the lack of semantic labels for many features, which complicates effective feature engineering and increases the risk of model overfitting. Given the accelerating adoption of cryptocurrencies and the evolution of criminal techniques for obfuscating transaction trails, there is an urgent need for research aimed at developing robust, scalable methods to detect illicit financial flows in anonymized blockchain environments.

Therefore, research on the development and application of automated feature selection methods based on genetic algorithms to improve the detection of illicit Bitcoin transactions is timely and highly relevant.

2. Literature review and problem statement

Detecting illicit transactions on the Bitcoin network has attracted significant scientific interest, leading to many studies focusing on various machine learning methods, feature selection strategies, and graph approaches.

A foundational study by [9] introduced the elliptic dataset, one of the first publicly available Bitcoin transaction graph datasets annotated with legality labels. The authors compared classical machine learning models, such as logistic regression, random forest, and multilayer perceptrons with graph convolutional networks (GCNs). Although GCNs did not achieve the highest F1-score (62.8%), they effectively captured topological dependencies in transaction graphs. Random Forest, however, outperformed both GCNs and logistic regression due to its ensemble voting mechanism, achieving higher predictive accuracy. The study also proposed ideas for combining graph neural networks with decision trees – for example, replacing the logistic regression output layer in GCNs with differentiable decision trees – but these approaches remain untested, highlighting a key direction for future research on integrating graph representations with ensemble learning.

The study [10] conducted a comprehensive evaluation of traditional and ensemble models, demonstrating that methods such as random forest, extra trees, and bagging significantly outperformed others in terms of accuracy (up to 98.13%) and F1-score (83.36%), affirming their effectiveness for anti-money laundering (AML) tasks. The authors showed that ensemble learning approaches were particularly successful in reducing false positives without increasing false negatives when applied to both local and aggregated transaction features derived from the Bitcoin graph. However, the work primarily focused on static classification performance without integrating more advanced graph-based deep learning techniques. The authors emphasized that effective preprocessing of graph-structured data remains a challenging problem and identified the exploration of graph structure-aware models and deep learning architectures as critical directions for future research.

The paper [11] presents the results of research on a hybrid architecture that combines Graph Convolutional Networks (GCNs) with linear layers and a Multi-Layer Perceptron (MLP) for predicting illicit transactions in the Bitcoin network. It was shown that this approach significantly outperforms standard GCN models on key metrics, achieving an accuracy of 89.9% and an F1 score of 77.3%. But there were unresolved issues related to the potential distortion of node representations caused by the spectral aggregation process and the limitation of GCNs designed primarily for undirected graphs. The reason for this may be that the normalization factors in the Laplacian matrix can unfairly weight the contribution of neighboring nodes, which complicates learning. A way to overcome these difficulties can be the concatenation of latent features from GCN layers with representations learned by linear layers, which was demonstrated to improve performance compared to using graph convolutions alone.

In the study [12], an extended dataset, Elliptic++, was developed to increase the number of features and labels to improve the classification performance of illicit transactions and accounts. The authors employed ensemble models, including Random Forest, XGBoost, and MLP, and applied feature selection based on feature importance, which enabled improvements in classification accuracy and recall. However, the study did not consider specialized feature selection methods adapted to the anonymous nature of blockchain data, such as evolutionary algorithms. Moreover, the potential for overfitting due to performing feature selection on the same dataset, as well as the lack of validation on other cryptocurrencies, raises questions about the transferability and robustness of the proposed approach.

The paper [13] presents the results of research on unsupervised approaches for detecting abnormal transaction behaviors in the Bitcoin network. It was shown that clustering and anomaly detection algorithms, such as K-means and other unsupervised models, can identify real cases of theft and fund loss without labeled data. But there were unresolved issues related to the lower precision and recall of these methods compared to supervised approaches. The reason for this may be the inherent difficulty of modeling pseudo-anonymous blockchain transactions and the absence of ground truth labels, which limits their practical effectiveness.

The paper [14] presents the results of research on active learning approaches to reduce the need for labeled samples in the classification of blockchain transactions. It was shown that combining active learning with supervised models can yield competitive results, especially in large-scale, partially labeled networks. But there were unresolved issues related to the poor performance of unsupervised methods compared to supervised algorithms. The reason for this may be the limited relevance of features and insufficient availability of high-quality labeled data.

In [15], the utility of random forest classifiers for measuring feature importance was examined, but without a rigorous selection process, redundancy in features can persist. This inefficiency is countered in [16], where heuristic optimization techniques, including particle swarm optimization and evolutionary computation, were employed for financial fraud detection. These studies demonstrate that metaheuristic algorithms like GAs not only improve classification metrics but also reduce training times, enhancing their viability for real-time applications. These generic methods, however, do not consider the anonymized structure of the Elliptic dataset, which poses unique challenges for interpretability and feature selection.

Finally, the study [17] applied genetic algorithms to feature selection in domains such as biometric security, demonstrating their capability to evolve high-quality feature subsets. It was shown that genetic algorithms can effectively reduce dimensionality while preserving predictive accuracy. However, their application in anonymized blockchain transaction networks has not been comprehensively investigated.

Summarizing the above studies, it can be noted that despite substantial progress in applying supervised, unsupervised, and ensemble learning methods to the detection of illicit transactions in the Bitcoin network, several critical limitations remain unresolved. These include the lack of specialized feature selection techniques adapted to the anonymous and high-dimensional nature of blockchain data, limited validation across different cryptocurrency networks, and insufficient integration of graph-based deep learning models with ensemble approaches. Given the accelerating adoption of cryptocurrencies and the evolution of criminal techniques for obfuscating transaction trails, all this suggests that it is advisable to conduct research focused on the development and application of automated feature selection methods based on genetic algorithms, tailored specifically to the specifics of anonymized Bitcoin transaction graphs. This approach is timely and highly relevant, as it has the potential to improve classification performance, enhance interpretability, and enable more efficient use of computational resources by reducing the dimensionality of the input data, while ensuring better generalization of predictive models and contributing to more reliable detection of illegal financial activities in decentralized systems.

3. The aim and objectives of the study

The aim of the study is to develop an evolutionary feature selection framework based on genetic algorithms for improving the classification of illicit Bitcoin transactions in anonymized blockchain datasets. This will enable more accurate, scalable, and interpretable identification of suspicious activity in cryptocurrency networks, where conventional manual feature engineering is limited due to data anonymity and complexity.

To achieve this aim, the following objectives were accomplished:

– to design and formalize the genetic algorithm-based feature selection process, covering data preparation, evaluation metrics, parameter configuration, and selection of optimal features;

– to implement the proposed method and experimentally assess its effectiveness by evaluating the impact of different parameter configurations on classification performance and computational efficiency;

– to compare the performance of the genetic algorithm based method with baseline models and state-of-the-art methods, using metrics such as precision, recall, and F1-score.

4. Materials and methods

4. 1. The object and hypothesis of the study

The object of this study is the process of classifying illicit Bitcoin transactions in blockchain datasets. The subject of the research is the application of feature selection methods based on genetic algorithms to improve the efficiency of classification models in this area.

The main research hypothesis is that applying a genetic algorithm-based feature selection method will improve the classification accuracy, precision, and recall of machine learning models by identifying a compact and informative subset of features.

The following assumptions are adopted in this study:

– the ground-truth labels in the Elliptic and Elliptic++ datasets are sufficiently reliable for supervised learning and evaluation;

– despite the anonymized nature of the features in the datasets, they contain enough structural information to support effective classification;

– tree-based and ensemble classifiers (e.g., Random Forest) are appropriate benchmarks for evaluating the impact of feature selection.

The following simplifications are adopted:

– the analysis is limited to the features provided in the Elliptic datasets, without incorporating external data sources or feature interpretation efforts;

– only classical and ensemble-based classifiers were used: deep learning or graph neural network architectures were not explored in this work;

– the study focuses on static data snapshots and does not include real-time detection.

4. 2. Theoretical methods

As described above, there are multiple machine learning algorithms that can be used for detection of illegal transactions. Through extensive research and review of relevant literature, it has been observed that Random Forest algorithm produces better results in various contexts compared to other supervised machine learning algorithms, as demonstrated in previous studies [9, 10, 12]. Random Forest is an ensemble learning technique consisting of multiple decision trees trained on random subsets of the dataset. It leverages the diversity of individual trees to enhance predictive accuracy and robustness. Each tree in the forest independently contributes to predictions, and the outcome is determined through a voting or averaging mechanism. Random Forest mitigates overfitting by introducing randomness in the tree-building process, such as selecting random subsets of features. Its application extends across diverse fields, demonstrating efficacy in handling complex relationships, reducing variance, and providing valuable insights into feature importance [18].

Considering the above advantages, it was decided to use Random Forest in this research as well. In addition to selecting the best algorithm, another important task to obtain a model with excellent accuracy is feature engineering or feature selection. Feature selection improves the performance of the model by focusing on the most important features and removing unnecessary or redundant features, which reduces the chance of overfitting and helps enhance model accuracy and making predictions more reliable.

In addition to the above-mentioned theoretical benefits, Feature selection also has several practical advantages. Firstly, it reduces computational complexity and costs. By identifying and storing only significant features, feature selection can significantly reduce the dimensionality of a dataset. Secondly, this dimensionality reduction, in turn, speeds up model training and reduces memory requirements. Third, it reduces the amount of disk space needed to store the data and the model. Fourth, increasing model efficiency means reducing processing time, making it especially suitable for real-time detection of illegal transactions.

In general, there are three main types of supervised feature selection methods [15]:

– filter methods: these methods assess the relevance of features based on their statistical properties and intrinsic characteristics, independent of the learning algorithm. Common techniques include correlation-based feature selection, mutual information, and statistical tests such as ANOVA or chi-square tests;

– wrapper methods: wrapper methods evaluate subsets of features by employing a specific machine learning algorithm to assess their performance. These methods typically use a search strategy to iteratively select subsets of features and evaluate them using the chosen learning algorithm. Examples include recursive feature elimination (RFE) and forward/backward feature selection;

– embedded methods: embedded methods integrate feature selection directly into the model training process. These techniques select features as part of the model building process, usually during the optimization of model parameters. Examples of embedded methods include Lasso (L1 regularization), decision tree-based feature selection).

To identify the most relevant features in this study, two methods were employed: the first is SelectKBest, which is a filter method, and the second is genetic algorithm, which is a wrapper method. More details about this will be in the experiments section. The metrics that were used to evaluate the performance of the models are the classic metrics: F1-Score, precision and recall. These metrics are widely used when dealing with imbalanced datasets, such as the dataset used in this study. Recall indicates the number of actual positive data the model was able to correctly predict. The mathematical equation of recall is shown in (1)

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{ False Negatives}}. \quad (1)$$

Precision is a metric that quantifies the number of correct positive predictions made. The mathematical equation of precision shown in (2)

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{ False Positives}}. \quad (2)$$

F1-score is calculated by the precision and recall value. The mathematical equation of F1-score shown in (3)

$$\text{F1 score} = 2 * \frac{\text{Precision*Recall}}{\text{Precision} + \text{Recall}}. \quad (3)$$

When faced with the task of identifying illegal transactions, it is the recall metric that is important, and therefore priority should be given to recall rather than precision, since the penalty for classifying an illegal transaction/account as legitimate far outweighs the penalty for the opposite [12]. Although the F1-score provides a balanced view of precision and recall, in the context of detecting illicit cryptocurrency transactions, recall remains the priority, as missing an illegal transaction (false negative) carries a higher risk than flagging a legal one. Weber [9] also noted that for anti-money laundering, more important is the illicit class.

4. 3. Datasets

During the study, 2 datasets were used: Elliptic and Elliptic++ bitcoin transaction datasets. The elliptic dataset associates Bitcoin transactions with real entities belonging to legitimate categories such as exchanges, wallet providers, miners, legitimate services, etc., not illegal ones such as fraud, malware, terrorist organizations, ransomware, Ponzi schemes, as well as unknown ones. The dataset includes about 200,000 transactions. The percentage breakdown is as follows: 2 percent were marked as illicit, 21 percent as licit, and 77 percent as unknown [9]. The Elliptic++ dataset, an extension of the Elliptic dataset that includes all Bitcoin wallet addresses and their temporal interactions related to transactions within the Elliptic dataset. By collecting Blockchain data associated with transactions, augmenting each transaction with 17 additional features, and compiling the dataset, the Elliptic++ dataset comprises over 822,000 labeled wallet addresses, each described by 56 features, and more than 1.27 million temporal occurrences [12]. Throughout this article, the term "Elliptic++" specifically refers to its transaction dataset. This convention will be maintained for clarity in all discussions regarding the Elliptic++ dataset.

4. 4. Experimental setup

After choosing a machine learning algorithm, an approach, datasets, and defining metrics to evaluate the model, the experiment was initiated. At first, all 166 features for the Elliptic dataset and 187 features for the Elliptic++ transaction dataset were used to train the model without any feature selection in order to establish baseline results. The objective was to compare these results with models in which feature selection methods would subsequently be applied.

The dataset was split into training and test sets in a 70:30 ratio. Specifically, the first 34 time steps were used for training the model, while the remaining 15 time steps were reserved for testing.

As an initial feature selection technique, SelectKBest from the scikit-learn library was applied. SelectKBest operates by scoring individual features using a specific statistical test to determine their importance and then selecting the top k features with the highest scores, where k is a user-defined parameter. Various k values were tested, but the results were worse than when using all features.

Subsequently, a genetic algorithm was applied for feature selection, as described in detail in the next subsection.

4. 5. Overview of genetic algorithms for feature selection

Genetic algorithms (GAs) are heuristic search algorithms within the broader category of evolutionary algorithms. They leverage the principles of natural selection and genetics to intelligently guide random searches using historical data, aiming to improve performance within the solution space. Widely applied in generating high-quality solutions for optimization and search problems, GAs mimic natural selection, favoring species that can adapt to environmental changes and reproduce. In essence, they emulate "survival of the fittest" among individuals across generations to address problems. Each generation comprises a population of individuals, with each individual representing a potential solution within the search space, often encoded as a string of characters, integers, floats, or bits [19].

4. 6. Computational environment

During this work, the experiments were conducted on an ASUS laptop, with Intel(R) Core (TM) i7-1065G7 CPU@ 1.30GHz 1.50 GHz based processor, 16.0 GB of RAM.

5. Results of applying genetic algorithm-based feature selection to illicit bitcoin transaction classification

5. 1. Design of the genetic algorithm-based feature selection framework

Fig. 1 illustrates the complete workflow of the genetic algorithm-based feature selection framework. It includes the following stages: data preparation, fitness evaluation, evolutionary search using genetic algorithm, selection of the best feature subset, and model training. Each of these steps is described below:

– data preparation: the dataset was divided into training and testing sets in a 70:30 ratio. Specifically, the first 34-time steps are used for training the model, and the remaining 15-time steps are used for testing;

– evaluation function: an evaluation function was defined to assess the quality of feature subsets. This function served as a guiding metric for the genetic algorithm, measuring the fitness of each candidate subset of features in relation to the classification objective;

– genetic algorithm: the core of this methodology is the genetic algorithm. Through successive generations, Genetic Algorithm iteratively refines subsets of features, aiming to maximize their performance in the context of scoring function. This selection method of choice, tournament selection (tools.selTournament), ensures a robust and diverse selection process, promoting the exploration of diverse feature combinations. Careful tuning of genetic algorithm parameters is paramount to the success of this methodology. Meticulously adjusting of parameters such as population size, number of generations, crossover probability, and mutation probability to strike a balance between exploration and exploitation within the search space. These parameters significantly influence the convergence and performance of the genetic algorithm;

– best individual: upon completion of all iterations, the genetic algorithm identifies the best-performing feature subset, termed the “Best Individual.” This subset represents the optimal configuration of features that achieved the highest fitness according to the evaluation criteria;

– final feature selection and model training: after selecting the optimal subset of features, a machine learning classifier was trained using the selected features to build a robust model. The classifier underwent an evaluation on the testing dataset to assess its performance and generalization capabilities. In the final evaluation phase, the model’s effectiveness in making accurate predictions on unseen data was systematically measured.

5. 2. Implementation and experimental evaluation on the public available datasets

Table 1 summarizes the optimal Genetic Algorithm configurations that achieved the best performance for each dataset, including the time required to select the optimal feature subsets. The results are reported in terms of precision, recall, and F1-score.

Table 1

The optimal Genetic Algorithm configurations and corresponding results

Dataset	GA parameters	Results	Time
Elliptic	Population = 250	Precision = 99.37%	41 hours
	Generation = 100	Recall = 73.13%	
	Crossover = 0.9	F1 = 84.26%	
	Mutation = 0.2		
Elliptic++	Population = 250	Precision = 99.37%	51 hours
	Generation = 100	Recall = 72.85%	
	Crossover = 0.9	F1 = 84.07%	
	Mutation = 0.2		

As can be seen from Table 1, for both datasets the best results were obtained with the same GA parameters: Population = 250, Generation = 100, Crossover = 0.9, Mutation = 0.2. But different amounts of time were spent on the GA to select features, since the Elliptic++ dataset has 17 additional features. T best model for Elliptic showed the following results: precision – 99.37%. recall – 73.13%, F1 Score – 84.25%. For Elliptic++ showed the following results: precision – 99.37%. recall – 72.85% and F1 Score – 84.07%. In terms of recall and f1-score the difference between the datasets were 0.18% and 0.28%, respectively, and in terms of precision they were the same.

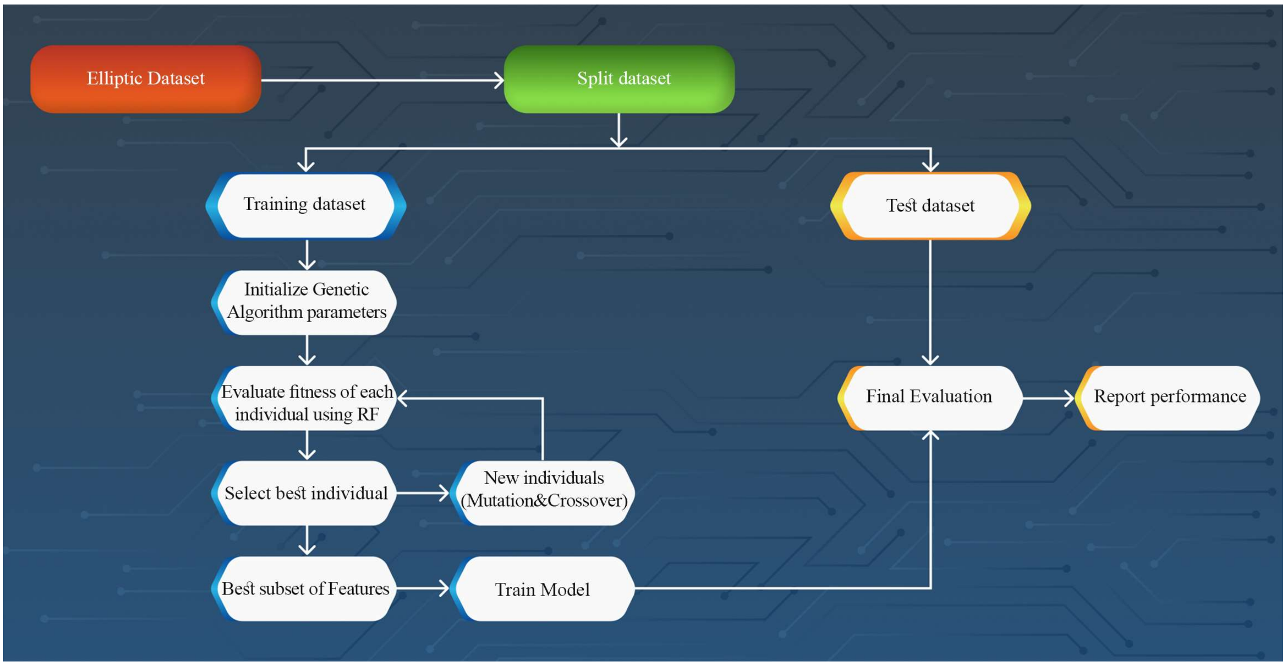

Fig. 1. Process of feature selection using GA and training model

5. 3. Performance comparison with baseline and state-of-the-art models

Table 2 and Fig 2. shows a comparison of the results of this study with those of similar studies using the Elliptic dataset. The results surpass those of comparative studies across all three metrics. Especially if look at the recall. [9] reported a result of 67%, while [10, 12] both achieved 71.9%; The result in this study, at 73.1%, exceeds these by 1.2%.

Table 2

Comparison with related studies using Elliptic

Reference	Precision	Recall	F1
[9]	95.6%	67.0%	78.8%
[12]	97.5%	71.9%	82.8%
[10]	99.1%	71.9%	82.4%
Current study	99.4%	73.1%	84.3%

Table 3 shows a comparison of the results of this study with those found in the work of [12] using the Elliptic++ transaction dataset, where it is also clear that this result outperforms theirs: 0.8% precision, 0.2% recall, and 0.5% F1 score.

Table 3

Comparison with the study by [12] using Elliptic++

Reference	Precision	Recall	F1
[12]	98.6%	72.7%	83.6%
Current study	99.4%	72.9%	84.1%

Furthermore, it is important to emphasize that the framework was evaluated on the complete Elliptic dataset and conducted experiments systematically increasing both the population size and the number of generations. It was observed that increasing these parameters in the genetic algorithm does not always guarantee better results. Excessive increases can lead to overfitting: the algorithm may begin to memorize training data instead of learning general patterns, resulting in poor performance on validation data. Moreover, larger population size and generation counts can slow down convergence or cause premature convergence, where the algorithm settles on a sub-optimal solution without exploring the entire search space.

For example, a configuration with a population size of 500, 250 generations, a crossover probability of 0.9, and a mutation probability of 0.1 resulted in 96.8% precision and 72.6% recall but required 989,462 seconds (approximately 11 days and 12 hours) of computation time., with no substantial improvement in recall compared to smaller configurations. This confirms that simply scaling up these parameters on the full dataset does not consistently yield better performance and, in practice, leads to diminishing returns and excessive resource consumption.

6. Discussion of the proposed genetic algorithm-based feature selection method

The integration of genetic algorithms (GA) for feature selection in the detection of illegal Bitcoin transactions has demonstrated considerable effectiveness in this study. By mimicking natural selection, GAs dynamically evolves feature subsets, ensuring that only the most informative characteristics are retained for classification tasks, as shown in Fig. 1. The high precision (99.4%) and improved recall (73.1%) and F1-score (84.3%) obtained in these experiments, as illustrated in Table 1, underscore the superiority of GA-based feature optimization compared to traditional approaches such as SelectKBest. This method addresses the challenge of identifying relevant features in anonymized blockchain data with many potential attributes, where semantic labels and explicit transaction characteristics are often unavailable. The capability of Genetic Algorithms to discover hidden structures in high-dimensional feature spaces without requiring manual annotation is particularly valuable in this context.

Compared to existing approaches described in [9–12], the proposed GA-based framework demonstrates several distinctive features. Firstly, it enables fully automated and adaptive identification of relevant features in anonymized blockchain data, which is not feasible with manual heuristics. Secondly, the method provides higher recall (+1.2%) and F1-score (+1.5%) relative to baseline implementations, as shown in Table 2 and in Fig. 2 for the Elliptic dataset.

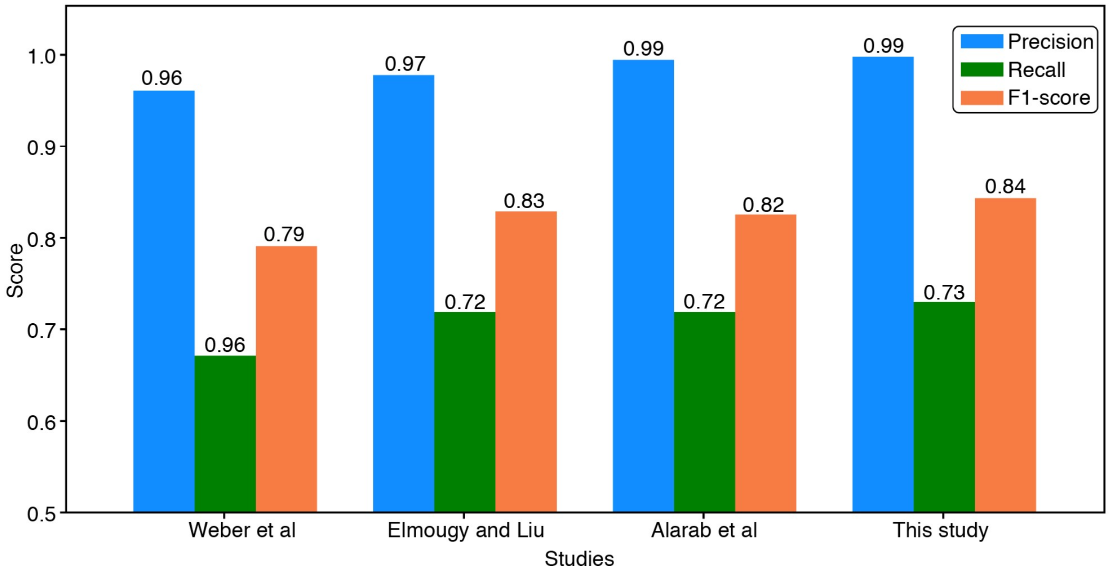

Fig. 2. Comparison of the results of four studies using Elliptic

Additionally, a comparison with the study by [12] on the Elliptic++ dataset is presented in Table 3. The results show that the proposed method achieves higher precision (99.4% vs. 98.6%), slightly higher recall (72.9% vs. 72.7%), and an improved F1-score (84.1% vs. 83.6%)

This finding aligns with recent research suggesting that heuristic search methods like GA significantly enhance detection accuracy when applied to complex and high-dimensional financial datasets. For instance, in [20], the combination of GA with random forest classifiers improved classification performance by isolating relevant financial indicators in a reduced feature space. Similarly, [21] demonstrated that gradient boosting-based models benefited from advanced feature engineering techniques, which not only improved accuracy but also decreased computational complexity, enabling faster model training and inference – an essential requirement for real-time fraud detection systems.

Importantly, these comparative experiments with both the Elliptic and Elliptic++ datasets confirm that optimal GA parameter tuning – specifically population size, generation count, crossover, and mutation probabilities configured as shown in Table 1 is critical to achieving robust feature selection.

The increasing complexity of illicit financial flows through blockchain platforms necessitates advanced analytical techniques capable of adapting to evolving criminal behavior. This study emphasizes that conventional static models are often insufficient to capture dynamic, multi-layered transaction structures within Bitcoin networks. By applying GA-driven feature selection, model demonstrates resilience in distinguishing illicit transactions despite class imbalance and temporal variability.

This is particularly relevant given recent trends in cryptocurrency misuse, as observed in [22], where market dynamics and decentralized structures were exploited to conduct illegal transfers. These activities often manifest through intricate transaction chains that evade detection by simplistic rule-based systems. Consequently, leveraging machine learning classifiers trained on GA-optimized features represents a promising direction for enhancing detection capabilities.

Furthermore, findings in this study support the argument presented in [12], which advocates for the adoption of ensemble learning strategies in money laundering detection. Ensemble methods, particularly when paired with optimized input features, improve generalizability across datasets and transaction types. In this case, random forest models trained on GA-selected features effectively mitigated overfitting and yielded stable performance across both datasets, thereby reinforcing their applicability to real-world cryptocurrency forensics.

In addition to accuracy metrics, recall values are critical in compliance-sensitive applications. The slightly improved recall observed in this study (73.1% for Elliptic) compared to the benchmark studies highlights the capacity of feature selection to enhance sensitivity toward illicit transactions, aligning with the assertion in [9, 12] that recall should be prioritized in financial crime detection to minimize false negatives.

Despite the positive outcomes, several limitations of this study should be acknowledged. One limitation lies in the computational cost associated with GA, particularly when applied to large, feature-rich datasets such as Elliptic++. As noted in [20], although heuristic optimization can yield superior results, it can also significantly increase training time and resource requirements, which may hinder applicability in low-resource or real-time environments. Moreover, the reliance on historical labeled datasets (Elliptic, Elliptic++) without external validation on alternative blockchain data may limit the generalizability of results across different platforms. The stochastic nature of GA introduces variability, potentially affecting reproducibility across multiple runs.

In addition to these limitations, certain disadvantages should be noted. First, the method requires manual configuration and tuning of GA parameters, which can be challenging for practitioners without prior experience in evolutionary algorithms. Second, even outside large-scale deployments, the approach demands access to high-performance computing resources to maintain reasonable processing times, making it less practical for organizations with limited infrastructure or for deployment on low-power or embedded devices. Future research could address these disadvantages by developing automated parameter optimization techniques and exploring lightweight implementations suitable for constrained environments.

Looking ahead, several opportunities and challenges for further development can be identified. While the model achieved robust performance on historical data, the adaptability of GA-based systems to real-time or streaming transaction scenarios remains to be explored. Research in [23] has proposed the use of LightGBM-powered systems for real-time blockchain analysis, highlighting the potential for incorporating more lightweight and adaptive algorithms in future implementations. There is also scope for integrating explainability mechanisms into detection pipelines. As emphasized in [24, 25], incorporating explainable AI (XAI) into financial crime models is vital for building trust and ensuring regulatory compliance. However, integrating explainability with heuristic optimization introduces methodological challenges, including the difficulty of maintaining consistent attribution of feature importance across evolving populations and ensuring the reproducibility of explanations over multiple GA runs. Additionally, the development of interpretable GA-based models may require substantial computational resources and careful tuning of explanation granularity to balance clarity and fidelity.

In summary, this study confirms the value of genetic algorithm-based feature selection in improving the classification of illegal Bitcoin transactions. The method effectively addresses the limitations of manual feature engineering and enhances model accuracy and recall. However, balancing predictive performance, interpretability, and computational feasibility remains a central focus for future research. Further work on optimizing resource efficiency, expanding cross-chain validation, and embedding explainability will be critical to realizing the full potential of this methodology in operational cryptocurrency compliance systems.

7. Conclusion

1. A robust evolutionary framework was developed that uses genetic algorithms to autonomously select meaningful feature subsets, eliminating the need for manual heuristics or prior domain knowledge. This framework is characterized by its consistent ability to produce stable and accurate classification results across different datasets, parameter configurations, and repeated experimental runs. Unlike common filter methods, the GA-based method leverages pop-

ulation evolution and fitness evaluation to explore non-linear feature interactions in anonymized blockchain data. In this context, the notion of robustness also refers to the reproducibility of the selected feature subsets and their effectiveness when varying algorithm parameters such as population size, crossover rate, and mutation probability. Such design choices make the framework particularly effective for the Elliptic dataset, where feature semantics are unknown, and traditional methods often fail to generalize. The resulting subsets significantly reduced dimensionality while maintaining high classification performance, demonstrating the framework's reliability and practical applicability in handling anonymized and high-dimensional inputs.

2. The framework was implemented and tested on both Elliptic and Elliptic++ datasets using random forest. The GA-selected feature subsets led to improved recall and F1-scores compared to models trained on full feature sets. For example, the best-performing classifier achieved an F1-score of 84.3%, precision of 99.4%, and recall of 73.1% on the Elliptic dataset. These results validate the adaptability of the framework across datasets and classifiers and demonstrate its practical utility for feature reduction in blockchain-based classification tasks.

3. Comparative analysis confirmed that the GA-based feature selection outperformed both baseline models using all features and the leading techniques previously applied to this task. Compared to the best published results, this method improved F1-score by 0.9%, precision by 0.3%, and recall by 1.2% on the Elliptic dataset. On Elliptic++, the improvements were similarly consistent. These performance gains are attributed to the GA's ability to optimize feature relevance in the absence of labeled feature descriptions, making it particularly effective for anonymized financial datasets where conventional feature engineering is infeasible.

Conflict of interest

The authors declare that they have no conflicts of interest in relation to the current study, including financial, personal, authorship, or any other, that could affect the study, as well as the results reported in this paper.

Financing

This research has been funded by the Science Committee of the Ministry of Education and Science of the Republic of Kazakhstan (Grant No. AP25793719).

Data availability

The manuscript has related data in the data repository. References are provided in the text of the paper.

Use of artificial intelligence

The authors confirm that they did not use artificial intelligence technologies when creating the current work.

References

1. Berentsen, A., Schar, F. (2018). The Case for Central Bank Electronic Money and the Non-case for Central Bank Cryptocurrencies. Review, 100 (2), 97–106. https://doi.org/10.20955/r.2018.97-106
2. Gajdek, S., Kozak, S. (2019). Bitcoin as an Electronic Payment Tool. Zeszyty Naukowe Uniwersytetu Przyrodniczo-Humanistycznego w Siedlcach. Seria: Administracja i Zarządzanie, 47 (120), 33–39. https://doi.org/10.34739/zn.2019.47.04
3. Bunjaku, F., Gjorgieva-Trajkovska, O., Kacarski, E. M. (2017). Cryptocurrencies – advantages and disadvantages. Journal of Economics, 2 (1), 31–39.
4. Sicignano, G. J. (2021). Money Laundering using Cryptocurrency: The Case of Bitcoin! Athens Journal of Law, 7 (2), 253–264. https://doi.org/10.30958/ajl.7-2-7
5. How terrorist groups are exploiting crypto to raise funds and evade detection. Elliptic. Available at: https://www.elliptic.co/blog/how-terrorist-organizations-are-exploiting-crypto-to-raise-funds-and-evade-detection Last accessed: 18.03.2024
6. Crypto crime mid-year update: Crime down 65% overall, but ransomware headed for huge year thanks to return of big game hunting (2023). Chainalysis Team. Available at: https://www.chainalysis.com/blog/crypto-crime-midyear-2023-update-ransomware-scams/ Last accessed: 18.03.2024
7. Crypto Crime Trends: Illicit Volumes Portend Record Year as On-Chain Crime Becomes Increasingly Diverse and Professionalized (2025). Chainalysis Team. Available at: https://www.chainalysis.com/blog/2025-crypto-crime-report-introduction/ Last accessed: 15.02.2025
8. Chuen, D. L. K. (2015). Handbook of Digital Currency: Bitcoin, Innovation, Financial Instruments, and Big Data. London: Academic Press. Available at: https://www.researchgate.net/publication/286223926_Handbook_of_Digital_Currency_Bitcoin_Innovation_Financial_Instruments_and_Big_Data Last accessed: 01.03.2024
9. Weber, M., Domeniconi, G., Chen, J., Weidele, D. K. I., Bellei, C., Robinson, T., Leiserson, C. E. (2019). Anti-Money Laundering in Bitcoin: Experimenting with Graph Convolutional Networks for Financial Forensics. Proceedings of the Workshop on Anomaly Detection in Finance (KDD '19). Anchorage. https://doi.org/10.48550/arXiv.1908.02591
10. Alarab, I., Prakoonwit, S., Nacer, M. I. (2020). Comparative Analysis Using Supervised Learning Methods for Anti-Money Laundering in Bitcoin. Proceedings of the International Conference on Machine Learning Technologies, 11–17. https://doi.org/10.1145/3409073.3409078

11. Alarab, I., Prakoonwit, S., Nacer, M. I. (2020). Competence of Graph Convolutional Networks for Anti-Money Laundering in Bitcoin Blockchain. Proceedings of the 2020 5th International Conference on Machine Learning Technologies, 23–27. https://doi.org/10.1145/3409073.3409080
12. Elmougy, Y., Liu, L. (2023). Demystifying Fraudulent Transactions and Illicit Nodes in the Bitcoin Network for Financial Forensics. Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. Long Beach, 3979–3990. https://doi.org/10.1145/3580305.3599803
13. Pham, T. B., Lee, S. (2016). Anomaly detection in bitcoin network using unsupervised learning methods. arXiv, arXiv:1611.03941. https://doi.org/10.48550/arXiv.1908.02591
14. Lorenz, J., Silva, M. I., Aparício, D., Ascensão, J. T., Bizarro, P. (2020). Machine learning methods to detect money laundering in the bitcoin blockchain in the presence of label scarcity. Proceedings of the First ACM International Conference on AI in Finance. New York, 1–8. https://doi.org/10.1145/3383455.3422549
15. Nerurkar, P., Bhirud, S., Patel, D., Ludinard, R., Busnel, Y., & Kumari, S. (2020). Supervised learning model for identifying illegal activities in Bitcoin. Applied Intelligence, 51 (6), 3824–3843. https://doi.org/10.1007/s10489-020-02048-w
16. Tayebi, M., El Kafhali, S. (2022). Performance analysis of metaheuristics based hyperparameters optimization for fraud transactions detection. Evolutionary Intelligence, 17 (2), 921–939. https://doi.org/10.1007/s12065-022-00764-5
17. Bouchlaghem, Y., Akhiat, Y., Amjad, S. (2022). Feature Selection: A Review and Comparative Study. E3S Web of Conferences, 351, 01046. https://doi.org/10.1051/e3sconf/202235101046
18. Breiman, L. (2001). Random Forests. Machine Learning, 45 (1), 5–32. https://doi.org/10.1023/a:1010933404324
19. Katoch, S., Chauhan, S. S., Kumar, V. (2020). A review on genetic algorithm: past, present, and future. Multimedia Tools and Applications, 80 (5), 8091–8126. https://doi.org/10.1007/s11042-020-10139-6
20. Contreras, R. C., Xavier da Silva, V. T., Xavier da Silva, I. T., Viana, M. S., Santos, F. L. dos, Zanin, R. B. et al. (2024). Genetic Algorithm for Feature Selection Applied to Financial Time Series Monotonicity Prediction: Experimental Cases in Cryptocurrencies and Brazilian Assets. Entropy, 26 (3), 177. https://doi.org/10.3390/e26030177
21. Taha, A. A., Malebary, S. J. (2020). An Intelligent Approach to Credit Card Fraud Detection Using an Optimized Light Gradient Boosting Machine. IEEE Access, 8, 25579–25587. https://doi.org/10.1109/access.2020.2971354
22. Howcroft, E. (2023). Crypto ransom attacks rise in first half of 2023, chainalysis says. Available at: https://www.reuters.com/technology/crypto-ransom-attacks-rise-first-half-2023-chainalysis-2023-07-12/ Last accessed: 15.02.2025
23. Aziz, R. M., Baluch, M. F., Patel, S., Ganie, A. H. (2022). LGBM: a machine learning approach for Ethereum fraud detection. International Journal of Information Technology, 14 (7), 3321–3331. https://doi.org/10.1007/s41870-022-00864-6
24. Kute, D. V., Pradhan, B., Shukla, N., Alamri, A. (2021). Deep Learning and Explainable Artificial Intelligence Techniques Applied for Detecting Money Laundering–A Critical Review. IEEE Access, 9, 82300–82317. https://doi.org/10.1109/access.2021.3086230
25. Černevičienė, J., Kabašinskas, A. (2024). Explainable artificial intelligence (XAI) in finance: a systematic literature review. Artificial Intelligence Review, 57 (8). https://doi.org/10.1007/s10462-024-10854-8