Tổng quan về ứng dụng mô hình Cây quyết định trong nhận diện gian lận giao dịch ngân hàng số

TÓM TẮT

Nghiên cứu thực hiện tổng quan có hệ thống về các nghiên cứu ứng dụng mô hình Cây quyết trong phát hiện gian lận trong giao dịch ngân hàng số, dựa trên dữ liệu thu thập từ các cơ sở Scopus, Web of Science và Google Scholar. Thông qua phương pháp phân tích trắc lượng thư mục và quy trình sàng lọc PRISMA, 76 bài báo phù hợp được xác định và phân tích chuyên sâu. Kết quả cho thấy các nghiên cứu hiện nay tập trung vào ba hướng chính: (i) hiệu suất của mô hình Cây Quyết Định trong việc phát hiện các gian lận trong giao dịch ngân hàng số; (ii) xử lý mất cân bằng dữ liệu bằng kỹ thuật SMOTE và các biến thể; và (iii) tối ưu hóa hiệu suất dự báo bằng cách tích hợp các phương pháp học máy khác tạo ra mô hình tổ hợp. Trên cơ sở đó, nghiên cứu đề xuất phát triển các mô hình Cây quyết định lai ghép có khả năng tự học và thích ứng, đồng thời kết hợp các kỹ thuật giải thích mô hình (XAI) như SHAP hoặc LIME để cân bằng giữa độ chính xác và tính minh bạch. Ngoài ra, cần tăng cường tích hợp dữ liệu thời gian thực và phân tích hành vi người dùng nhằm xây dựng hệ thống phát hiện gian lận thông minh, bền vững và đáng tin cậy hơn trong tương lai.

Từ khóa: *Cây quyết định, Ngân hàng số, Phát hiện gian lận giao dịch, Phương pháp Prisma*

A Review of Decision Tree Applications in Digital Banking Transaction Fraud Detection

ABSTRACT

The study conducted a systematic review of the research on applications of Decision Tree models in detecting fraud in digital banking transactions, based on data collected from Scopus, Web of Science, and Google Scholar. Through bibliometric analysis and the PRISMA screening process, 76 relevant articles were identified and analyzed in depth. The results showed that current research focuses on three main directions: (i) the performance of Decision Tree models in detecting fraud in digital banking transactions; (ii) handling data imbalance using the SMOTE technique and variants; and (iii) optimizing the forecasting performance by integrating other machine learning methods to create ensemble models. On that basis, the study proposed to develop hybrid Decision Tree models with self-learning and adaptive capabilities, and combine model explanation techniques (XAI) such as SHAP or LIME to balance between accuracy and transparency. In addition, it is necessary to strengthen the integration of real-time data and user behavior analysis to build a more intelligent, sustainable, and reliable fraud detection system in the future.

Keywords: *Decision Tree, Digital Banking, Transaction Fraud Detection, PRISMA.*

1. INTRODUCTION

Global digital transformation is an inevitable trend as technology is booming strongly. All sectors and industries in the economy are competing to invest in technology to improve services, enhance customer experience, to enhance competitive position, towards sustainable development. As digital banking develops strongly, along with the trend of cashless payments, more and more customers prefer to transact through bank apps rather than transact directly at the counter, this has posed a huge challenge in terms of security as well as ensuring network security for banks because strong technology development is also the main cause of the increase in fraudulent behavior as well as online fraud. According to an article published in 2025 by Thanh Luan in Thanh Nien newspaper, up to 2.3 million bank cards in Vietnam had their information leaked on dark websites in 2023 and 2024; however, up to 95% of the cards are still valid, so criminals can still easily carry out illegal acts on customers' accounts without being detected. In addition, in the report of Viettel Cyber Security Company on the situation of user account data leakage in Vietnam in 2024, there were about 14.5 million user accounts with personal information leaked, including social network accounts such as Facebook, Zalo, e-wallet accounts, email accounts, bank accounts, accounting for 12% of the total number of accounts with information leaked globally, which has increased the risk of bank accounts being hacked and easily stolen. If in the past, traditional fraud was simply finding passwords to log into customers' accounts or through fake messages or links asking customers to enter codes to steal information, now, technology criminals can simulate user behavior by using artificial intelligence to bypass authentication and log into customer accounts to stole money, this causes great obstacles to the control and management of banking transaction risks.

Faced with the increasing challenge of fraud, applying artificial intelligence (AI) to building a fraud warning and prevention system is urgent and practical. According to an article on SecurityBrief Asia, about 70% of global financial institutions have deployed AI and machine learning fraud detection systems in 2024. When AI is strongly applied in the banking and finance sector, supervised machine learning is considered an effective approach in identifying and predicting transactions with signs of fraud based on historical data sets. Among the commonly applied supervised machine learning algorithms, the Decision Tree model stands out with its ability to build classification rules in an intuitive and easy-to-understand way, and can analyze in-depth characteristics to identify fraudulent behavior quickly and easily. In particular, Decision Tree is considered a form of Explainable AI - XAI, because this model clearly shows the relationship between input variables and

predicted results through a series of specific decision rules. This explainability not only helps users understand the reasoning behind each fraud prediction or warning, but also meets the requirements of transparency and accountability in the banking sector.

In the context of increasingly sophisticated forms of financial fraud and credit risk, reviewing studies applying the Decision Tree model in detecting fraud in digital banking transactions not only helps guide future research on the application of the Decision Tree model in the banking sector but also provides practical implications for banks in selecting, deploying, and optimizing explainable artificial intelligence systems.

2. LITERATURE REVIEW

2.1. Transaction fraud detection

According to research by Vanini et al., online banking fraud occurs when criminals hijack and transfer money from an individual's online bank account.[1] E-banking fraud is costing billions of dollars globally, with cases such as phishing and identity theft leading to the loss of money from personal and business accounts.[2] These losses not only directly affect users but also increase systemic risks for banks and financial institutions, requiring increasingly effective prevention and detection solutions.

In the digital banking context, each transaction is conducted through a highly digitized and automated system, making fraud detection not only dependent on manual controls but also requiring intelligent data analytics models that can recognize unusual behavioral patterns in the transaction flow. Unlike traditional fraud, online transaction fraud occurs in real time and can be hidden among millions of legitimate transactions, making it difficult for early warning systems without effective detection models. Fraud is becoming more sophisticated through the use of artificial intelligence, such as deepfakes and fake calls, with tricks such as refund scams and QR codes.[3] Similarly, the Bank for International Settlements report also pointed out that fraudsters use malicious codes to automate fraudulent transactions, intercept authentication messages, and change recipient information, which are "dynamic" tools to deal with new banking technology.[4]

It can be seen that fraudulent transactions in digital banking are increasingly sophisticated and unpredictable, while increasing systemic risks for banks and financial institutions. These losses not only directly affect customers but also require increasingly intelligent and timely prevention and detection measures. Traditional monitoring systems and manual controls are no longer effective enough, especially when fraudulent transactions can be hidden among millions of valid transactions. The banking industry report has highlighted that banks should invest in AI, machine learning, and behavioral analytics to detect fraud in real-time, especially against threats from AI genes.[7] This shows that the application of smart technologies is essential for banks to proactively detect unusual behavior and improve the ability to protect customer assets and data. The Vanini et al. research group emphasizes digital banking fraud focused on online and mobile payment channels, in which fraudsters use identity theft to access the system as legitimate account holders, but their transaction behavior is different from the account holders.[5] Thus, it is possible that due to the differences in transaction behavior, fraud detection systems can detect early that a customer's account is being compromised, and take preventive actions such as locking the account and sending warnings to customers. Thus, transaction behavior analysis becomes a key tool to identify abnormalities early, helping the system respond promptly with measures such as locking accounts or sending warnings to customers, thereby minimizing the risk of financial loss for both banks and customers.

In the study by Wei et al., the proposed process for identifying fraudulent transactions using machine learning is as follows:[6]

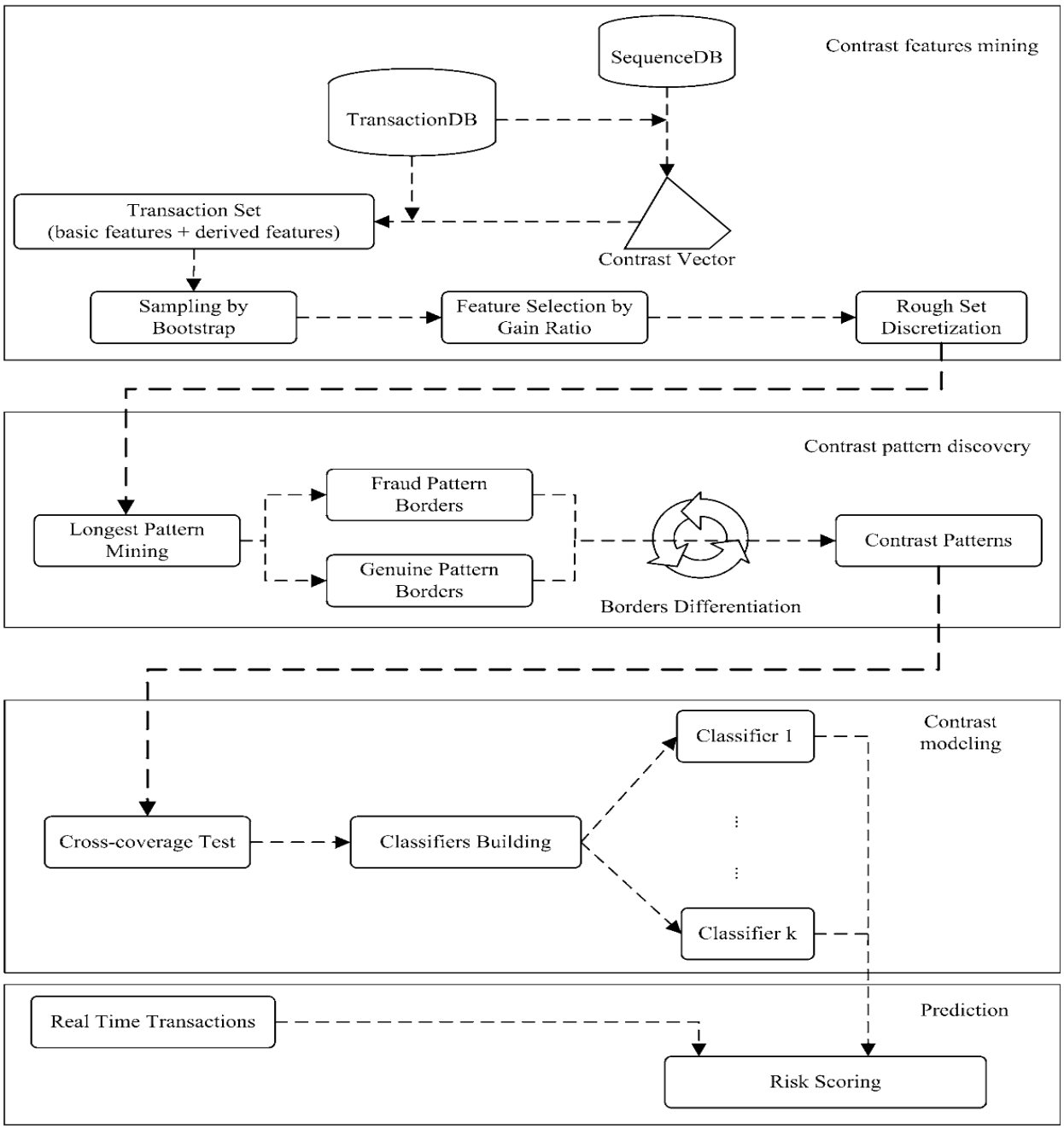

Figure 1. Framework for mining contrast in online banking behavior

First, the original transaction data is collected from the banking system and pre-processed to remove noise, standardize the format, and handle missing values to ensure the accuracy and integrity of the data. From this data set, features are extracted, including basic features such as transaction value, transaction type, execution time, and behavioral features inferred from the user's activity chain, reflecting real-time transaction habits and trends. After extraction, the data is put into the feature selection stage using information criteria to identify factors that have a high ability to distinguish between valid and fraudulent transactions. These features are further discretized to facilitate pattern mining and reduce computational complexity. Next, the system proceeds to mine contrasting behavioral patterns between the two groups of transactions, thereby detecting characteristic patterns that represent fraudulent behavior. Based on these patterns, a machine learning model is built through classifiers, which helps identify and group transactions according to their risk level.

In the final stage, the model is applied to real-time transactions to calculate the risk score for each transaction. Transactions with high risk scores will be alerted by the system for timely review and handling by the bank, contributing to

minimizing losses and improving fraud control efficiency.

The close coordination between these steps helps the system not only detect abnormal behavior but also continuously learn and adapt to new forms of fraud in the increasingly complex digital transaction environment.

In short, transaction fraud in digital banking is the most sophisticated, dynamic, and difficult to detect form of fraud in the modern financial environment. As artificial intelligence continues to grow, especially in the field of machine learning with its strengths in predicting and classifying objects, the research and application of machine learning models are becoming an effective direction in fraud detection and fraud risk management. “Machine learning algorithms and high processing power increase the capability of handling large datasets and fraud detection in a more efficient manner,” according to Hashemi et al.[7]

2.2. Decision Tree algorithm

A decision tree is a supervised machine learning model used to classify or predict based on the logical branching structure of data. According to Quinlan's research, a decision tree works on the principle of sequentially splitting data into smaller groups, maximizing the information purity of each group through measures such as Information Gain.[8] On the online learning platform GeeksforGeeks, the decision tree is visually drawn with each node in the tree representing an attribute, the branch representing the splitting condition, and the leaf node representing the predicted outcome label, as shown in Figure 2:

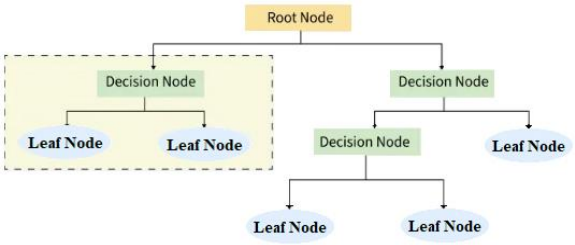

Figure 2. Basic structure of a Decision Tree

This model generates interpretable decision rules that help identify unusual behavioral patterns in digital banking transaction data, such as unauthorized transactions or suspicious logins. With high transparency, decision trees belong to the group of explainable artificial intelligence, giving them an advantage over complex models such as deep neural networks.

In the digital banking domain, fraudulent behavior is often demonstrated by unusual patterns in transaction behavior, such as sudden increases in transaction frequency, device changes, or unusual login locations. Decision trees help classify these patterns into “normal” and “fraudulent” groups based on threshold values learned from training data. A Decision Tree is a basic model in data mining for fraud detection, with good performance in classifying fraudulent transactions.[9]

According to bibliometric analysis results, since 2010, research on supervised machine learning applications, especially Decision Trees in fraud detection, has been a trend that many researchers are interested in.

Figure 3. Word cloud

Figure 3 shows a keyword cloud illustrating popular research topics related to fraud detection and the application of artificial intelligence in financial crime prevention during the period 2010 - 2020. Large keywords such as “decision trees,” “machine learning,” “fraud detection,” “crime,” and “random forests” show a high frequency of occurrence and great interest in this field. From this, it can be seen that the research trend of scientists on financial fraud detection focuses mainly on the application of machine learning methods, especially decision-making. Wei et al. asserted that “Decision tree is widely used in analysis, and the rules generated by the Decision Tree are easy to understand.”.[6] In summary, the Decision Tree algorithm plays a core role in digital banking fraud detection systems due to its ability to combine high predictive efficiency and interpretability, so it deserves to be exploited in depth.

3. METHOD

The study uses two methods, including bibliometric analysis and systematic review according to the PRISMA framework. These methods help provide a comprehensive overview of research development, key themes, and knowledge gaps in the selected field.

In the first phase, the study performed the bibliometric analysis. Scopus was chosen as the main data source because it is the world's largest and most prestigious academic database, with a standard, uniform structure, and supports exporting raw data files suitable for processing with bibliometric analysis tools using Bibliometrix software. Bibliometric analysis helps identify research trends, as well as prominent topics related to the Decision Tree model in the banking sector, thereby suggesting future research directions.

In this study, a dataset of 1,206 studies related to supervised machine learning applications to detect fraud in the financial and banking sector in the period 2010 - 2024 was collected from the Scopus database. The collected data were processed and analyzed using Bibliometrix software (R package) to identify research trends of scientists in the period 2010 - 2024 associated with the Fourth Industrial Revolution, when new technologies such as artificial intelligence, Internet of Things (IoT), big data, and cloud computing developed strongly and are widely applied.

Table 1. Number of published studies

Year	Articles
2010	17
2011	18
2012	20
2013	9
2014	12
2015	15
2016	29
2017	30
2018	64
2019	89
2020	133
2021	158
2022	176
2023	176
2024	260

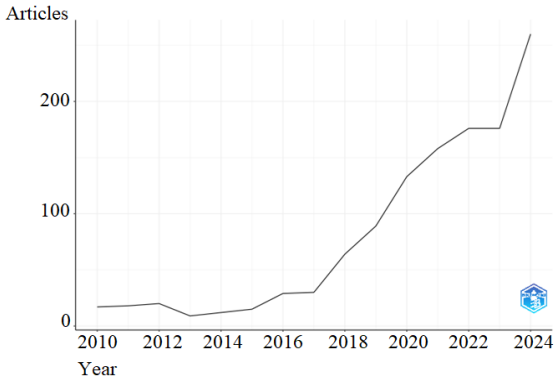

Figure 4. Chart of annual scientific production

It can be seen that the trend of research on supervised machine learning applications in the banking and finance sector has increased over the years, especially in the period 2020 - 2024, from 133 studies to 260 studies. With the strong development of technology and global digitalization, the application of artificial intelligence in the banking sector is extremely necessary, which is also the reason why research on machine learning applications in banking is prioritized by researchers.

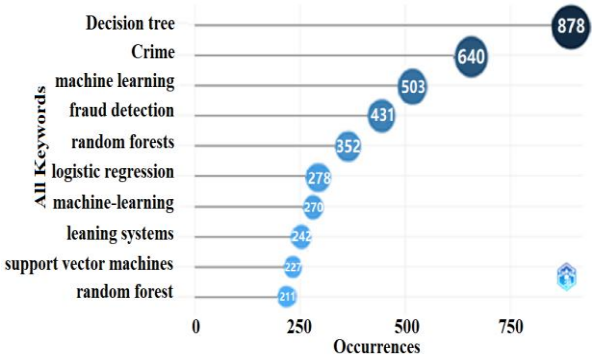

Figure 5. Frequency of keywords

Figure 5 shows the frequency of keywords in 1,206 studies on supervised machine learning applications in the banking sector from 2010 to 2024. Groups of phrases such as "Decision tree", "machine learning", "random forest" are model terms that are mentioned a lot in studies, especially "Decision tree" with 878 times. The phrases "crime" and "fraud detection" which are related to the application field of supervised machine learning appear quite a lot with frequencies of 640 and 431 times respectively, showing that these are two application fields of great interest. The third group of phrases, belonging to the group of classical methods, such as "logistic regression" and "support vector machines" also appear in studies but much less than "Decision tree" and "random forest", proving that these two methods are less popular.

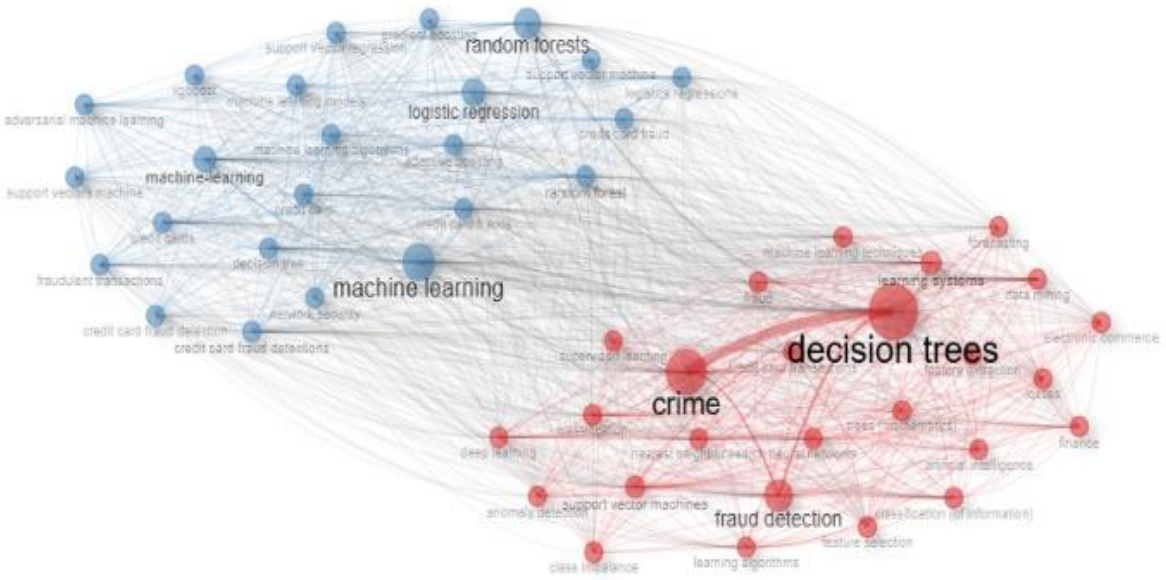

Figure 6. Keyword network

From Figure 6, it can be seen that in the period 2010 - 2020, the keyword network appeared with two clearly separated clusters, including the red cluster with the prominent keywords "Decision trees" and "Crime", while the blue cluster with the prominent keywords "machine learning" and "random forests".

The red cluster shows the application direction of machine learning in the banking and finance sector, especially the application of "decision trees" in identifying fraud and crime, shown through the association with the keywords "fraud detection", "crime", "forecasting", and "feature selection".

The blue cluster with the central keywords “machine learning”, “random forests”, “logistic regression”, and “xgboost” is the foundational machine learning models, along with the learning methods including “adversarial” and “supervised”, with the application directions of “fraudulent transactions”, “credit card fraud detection”. This group represents the theoretical and technical foundations in developing and improving machine learning algorithms.

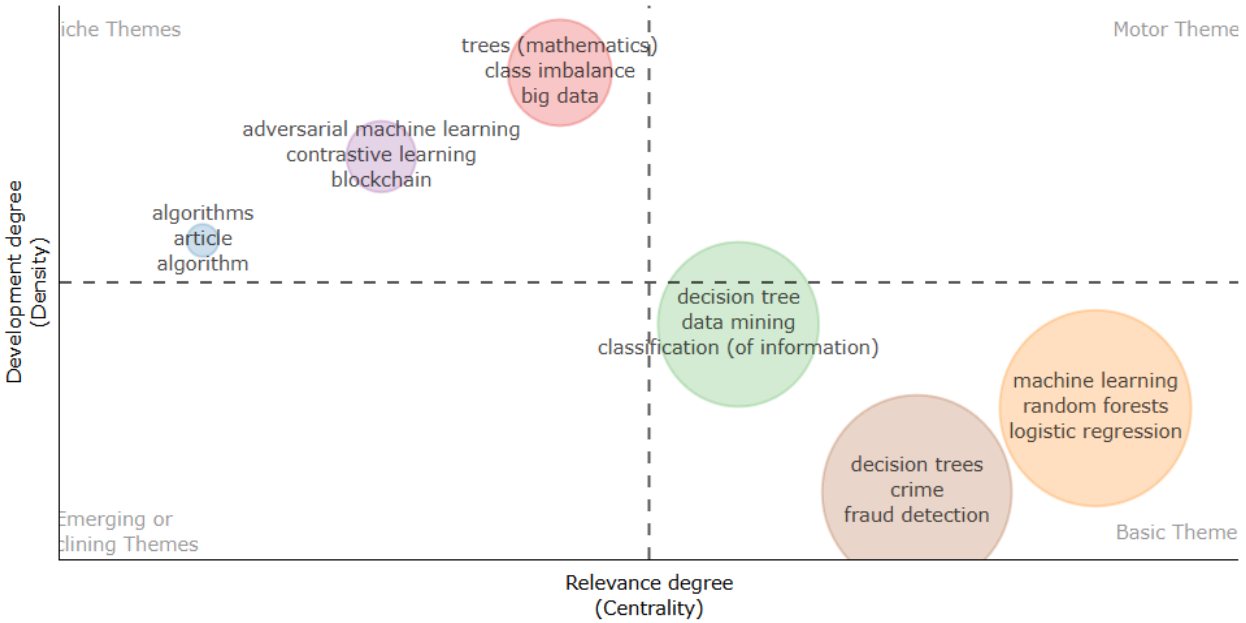

Figure 7. Thematic Map

Figure 7 shows four topic groups, including Niche Themes, Motor Theme, Basic Theme, and Emerging or Declining Theme. The chart shows the relationship between the level of development and the relevance of research topic clusters in the field of supervised machine learning. Each bubble represents a keyword phrase, in which the bubble size represents the importance or the number of related research.

The first is the Nick Theme group, which includes topics such as “trees,” “class imbalance,” “big data,” “adversarial machine learning,” “contrastive learning,” and “blockchain.” This group focuses on advanced machine learning methods and large-scale data processing, and extends to modern technologies such as blockchain and adversarial machine learning. Topics in this group are often researched at a deep level, but their scope of application is limited. Notably, the smallest ball is located at the boundary between Nick Theme and Emerging or Declining Themes, representing topics about “algorithms” that are general, have low density of development, and low level of connection, indicating a gradual decline in the future.

Next is the Basic Themes group with topics on “machine learning”, “random forests”, “logistic regression”, “decision tree”, which are highly central, play a fundamental role in the field of machine learning, and are commonly used in applications; however, the level of development of these topics is not deep. In particular, the “decision tree” ball is located at the intersection between Basic Theme and Motor Theme, showing that this topic is both fundamental and has a strong development and application trend. This reflects that Decision Tree not only plays a core role in traditional machine learning models, but also is an important bridge between theory and advanced applied research directions, especially in fraud detection, data mining, and trend prediction. In addition, the central position and intersectional behavior prediction. This combination not only increases accuracy but also improves the ability to handle large and complex data in the digital banking environment. Thus, the Basic Themes group is not only a theoretical foundation but also a hub connecting applied research, playing a key role in promoting the development of machine learning methods in practice. The role of the Decision Tree also shows its importance in integrating with other advanced techniques, such as ensemble learning or hybrid models, to improve the efficiency of fraud detection and

In summary, through bibliometric analysis on a dataset of 1,206 research articles on supervised machine learning applications in fraud detection in the financial and banking sector, it can be seen that in the period 2010 - 2024, scientists paid special attention to exploiting the combination of foundational models and modern specialized techniques. Foundational supervised machine learning models such as random forests, logistic regression, and decision trees continue to play an important role, with the Decision tree standing out as the core model, with a tendency to develop strongly and be widely applied, thanks to its intersection between the two areas of Basic Theme and Motor Theme. This clearly reflects the current research trend: both maintaining a solid theoretical foundation and expanding advanced application directions, especially in fraud detection systems and financial risk prediction.

In the systematic review phase, according to the PRISMA framework, studies were collected from reputable academic databases such as Web of Science, ScienceDirect, and Google Scholar, to expand the scope as well as increase the diversity of research areas and contexts. At the same time, from the 1,206 studies used in the bibliometric analysis, the research on the application of supervised machine learning to identify financial fraud is quite wide, so at this stage, we will screen and re-select works focusing on the Decision Tree, including both single and optimized models in fraud identification in digital banking. This approach not only ensures the reliability and representativeness of the data but also reflects more comprehensively the trends, results, and research directions in many different contexts related to the application of the Decision Tree in fraud detection and banking risk management. The PRISMA flowchart was developed by Page et al. to illustrate the process of identifying, screening, evaluating, and selecting studies in the systematic review process.[10] Studies are identified through various sources, and duplicate or ineligible studies are subsequently eliminated. The remaining studies are then assessed for relevance and selected again before inclusion in the overall analysis. This process helps ensure transparency, focus, and comprehensiveness in document selection. The steps in the process are illustrated in Figure 8:

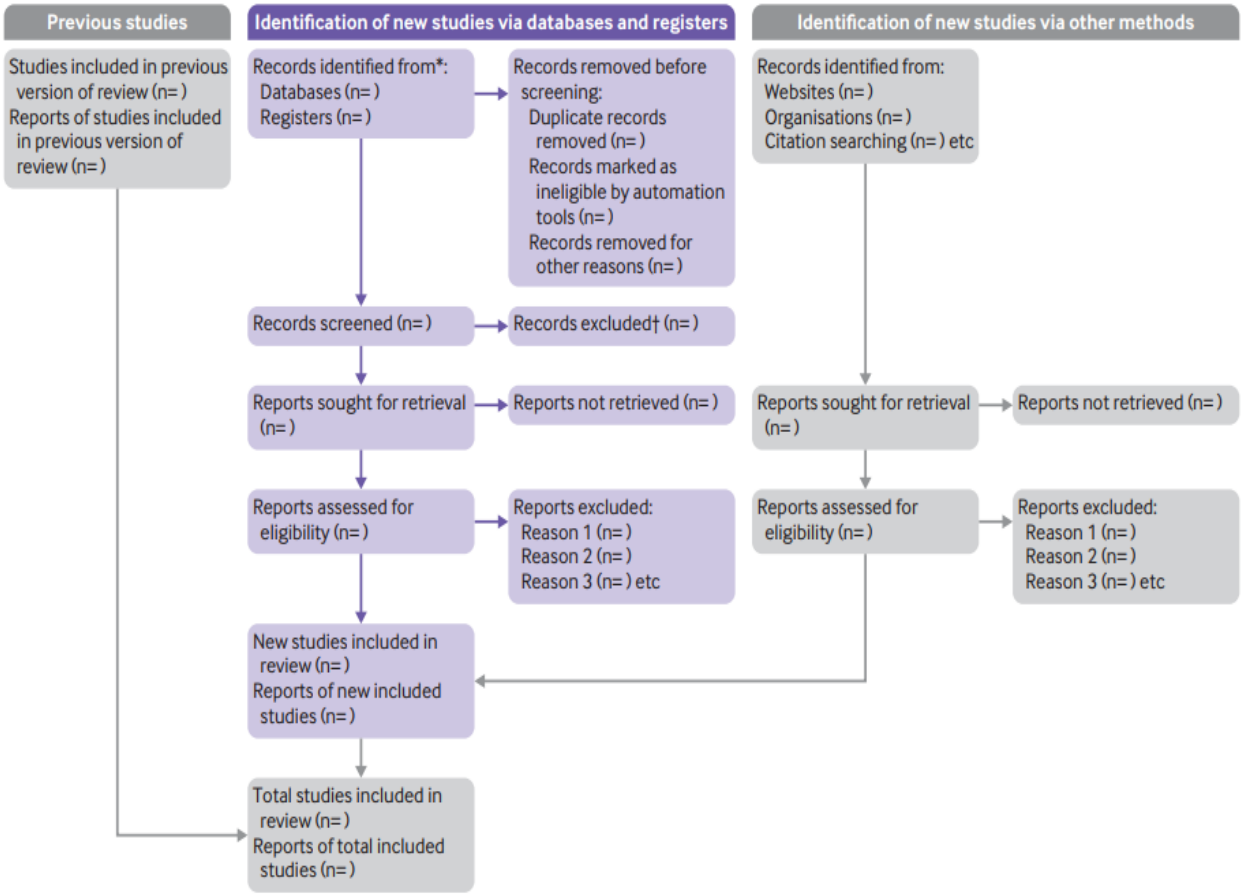

Figure 8. PRISMA 2020 flow diagram template for systematic reviews

First, 1,206 research articles on supervised machine learning applications in financial fraud detection were selected from Scopus in the bibliometric analysis section along with 238 studies identified from reputable academic databases such as Web of Science, ScienceDirect and Google Scholar, with search keywords related to "Decision Tree", "Fraud Detection", "Financial Fraud" and "Digital Banking" were screened and duplicates, short conference papers, or articles without full text were removed, leaving 671 articles for title and abstract screening to assess their preliminary relevance.

In the title and abstract screening stage, the evaluation process focused on determining the preliminary relevance of the 671 research articles to the research objectives. First, studies in the banking and finance field that are not related to fraud detection or financial fraud risk management will be excluded, including studies that focus on customer classification, credit prediction, or credit scoring but do not mention fraud factors. Second, studies on supervised machine learning that do not use the Decision Tree model as the main method, such as studies that use other algorithms as the main model, without analyzing the effectiveness of the Decision Tree. After this evaluation process, a significant number of papers were excluded because the title and abstract did not clearly show the connection between the Decision Tree model and the goal of fraud detection in digital banking, or only described the general application in financial data analysis without the risk identification aspect. The number of studies left was only 412 papers. Next, the author continued to evaluate the full text of 412 research articles with three exclusion criteria applied, including:

(i) The study does not focus on the topic of detecting fraud in digital banking transactions.

(ii) The Decision Tree model is not the main model or is not directly related to fraud detection.

(iii) The article lacks information about data or model results.

At this stage, each article was read and evaluated in detail regarding the scope of the study, model structure, experimental data and quantitative results. Articles that only presented theory, did not have simulations, or did not provide enough information to compare the model performance were eliminated. In addition, studies that only used Decision Trees as part of the preprocessing or feature selection process without conducting an independent evaluation of fraud detection performance were also not included in the final list. In addition, works with data samples of unknown origin, lacking performance evaluation indicators such as Accuracy, Precision, Recall, F1-score, or AUC, or only describing the overview concept without specific experimental results were also excluded from the final set.

After completing the full-text screening step, only studies that fully met the criteria, including the application of a single Decision Tree model or a combined model in which the Decision Tree is the underlying model applied in detecting and predicting fraud in digital banking transactions, were retained for the synthesis and analysis stage. As a result, 76 research articles were in line with the objectives of this study, eligible to be included in the classification step according to the research objectives, applied models, and experimental results.

4. FINDINGS AND DISCUSSION

To synthesize the insights from the 76 reviewed studies, Table 2 presents a summary of the key findings related to the application of Decision Tree models in detecting fraudulent transactions in digital banking. The table highlights the main methodologies used, their advantages, inherent limitations, and suggested improvement strategies, while also providing a brief overview of current research trends and practical considerations in the field.

Table 2. Summary of key findings

No.	Key findings	Methodology / Model	Advantages	Limitations	Improvement Methods
1	Effectiveness of the Decision Tree in detecting fraudulent transactions in digital banking	Single Decision Tree Model	Easy to interpret, transparent, simple, fast data processing	Easily overfitted; low accuracy when the dataset is imbalanced or unstable	Apply pruning, tune parameters, combine with machine learning models, and handle imbalanced data
2	Addressing the data imbalance problem	SMOTE and its variants	Improves the detection of minority classes	May generate noise or non-representative data; inappropriate sampling ratios may lead to overfitting	Combine resampling with cost-sensitive learning or optimize SMOTE parameters using Grid Search or Bayesian Optimization
3	Enhancing the prediction accuracy of fraud detection	Hybrid or ensemble models based on the Decision Tree	Good generalization ability, stable, less sensitive to noise	Time-consuming training process; model performance depends on the ensemble technique used	Analyze feature importance, select suitable ensemble methods, and design generalization-oriented models

The first important finding is that most studies confirm that the Decision Tree is one of the most popular and basic supervised machine learning models, often used as a foundation model in data classification problems based on simple, clear, and easy-to-interpret logical rules.[11-13] In many studies in the past, the authors pointed out that the outstanding advantages of this model lie in its transparency, intuitive explainability, and fast processing speed, making it a useful tool in problems that require quick and accurate decision making, especially in the financial sector, where high requirements for model explainability and auditability are necessary.[14-16] Decision Tree is often applied to detect fraudulent transactions, classify customers, or rate credit risk, thanks to the tree

structure that clearly shows the relationship between features and classification results. However, besides these advantages, the Decision Tree also reveals certain limitations. One of the biggest disadvantages is that it is susceptible to overfitting, especially when the model is too complex or when the training data is noisy, leading to poor generalization ability when applied to new data. In addition, this model is also sensitive to data imbalance – a phenomenon that is quite common in the digital banking field, where fraudulent transactions account for only a very small proportion of the total transactions. In studies using real banking data, authors all noted a significant decline in the performance of the Decision Tree in data sets with a high class mismatch ratio.[6,17-19] In response to this situation, many improvements have been proposed. The authors recommend pruning the tree to reduce depth and remove unnecessary nodes, helping the model avoid overfitting; at the same time, optimizing node splitting parameters such as maximum depth, minimum number of samples at each node, and so on to achieve a balance between accuracy and generalization ability. In addition, combining Decision Tree with ensemble models such as Random Forest, Gradient Boosting, or hybrid models combined with LSTM, Isolation Forest, or AI-powered systems has been proven to be more effective in increasing the stability and generalization ability of the model.[11,14,15,20,21]

In addition, many other studies also focus on handling imbalanced data before training, through techniques such as oversampling, undersampling, or feature engineering, to improve the accuracy in identifying rare patterns and improve the overall efficiency of the fraud detection system.[6,22]

The second important finding relates to the ability to handle data imbalance, one of the key challenges in fraud detection. In practice, the proportion of fraudulent transactions is usually tiny, only about 0.1–1% of the total number of transactions, making machine learning models, including Decision Trees, susceptible to majority bias. As a result, the model can achieve high overall accuracy but has low sensitivity to rare fraud cases. Some research clearly pointed out this limitation, emphasizing that improving minority detection is a core factor to increase the effectiveness of anti-fraud systems in banks.[13,23] To address this issue, many works have proposed the application of SMOTE and its variants, such as K-SMOTEENN, SMOTE-Tomek Links, or SMOTE combined with reinforcement learning. Other studies have demonstrated that generating additional artificial samples from the minority class feature space helps balance the data distribution, expand the decision boundary, and significantly improve the ability to detect fraud.[24-26] In addition, combining SMOTE with enhanced tree algorithms such as Random Forest, XGBoost, or Gradient Boosting is also noted to help increase sensitivity and F1-score, minimizing data bias.[27,28] However, if not applied properly, SMOTE can generate synthetic data points that do not reflect the real-world distribution, thereby distorting the characteristics of the training data and making the model susceptible to overfitting.[29] Therefore, many new approaches propose combining SMOTE with cost-sensitive learning, in which the model is assigned a higher weight to the errors of the minority class, providing a better balance between accuracy and detection ability. Findings from a previous study show that combining SMOTE and cost-sensitive learning with parameter optimization using Grid Search or Bayesian Optimization has increased the stability of the model and improved the ability to accurately detect rare fraudulent transactions in today's digital banking environment.[26]

The final finding shows that the prominent trend of recent studies is to improve the accuracy and stability of fraud prediction models through the combination of machine learning models, especially advanced tree models such as Decision Tree, Random Forest, Gradient Boosting, or XGBoost.[26,30-33]. These methods take advantage of combining multiple decision trees, which helps increase the ability to resist noise, reduce prediction errors, and improve the generalization ability of the model. However, the significant limitations are high complexity, long training time, and difficulty in interpreting results - factors that are especially important in the financial sector, which requires transparency and clear model explainability. To overcome this, recent studies have proposed many optimization directions, such as feature importance analysis, removing ineffective features, or integrating model explanation methods such as SHAP and LIME to clarify the operating mechanism of the model, ensuring both maintaining high accuracy and increasing transparency in decision making.[32,34-36]

Synthesizing previous research results shows that Decision Tree still plays a key role in fraud detection and risk management in digital banking. Although financial transactions are

increasingly complex and diverse, this model is still considered the foundation thanks to its clear structure, high interpretability, and flexibility in combination with other models. However, the effectiveness of the Decision Tree depends largely on the degree of parameter optimization, the quality of input data, and the ability to integrate with complementary machine learning methods.

In that context, the current research trend is strongly shifting from simple decision tree models to hybrid or ensemble models to both improve accuracy and maintain explainability. At the same time, the integration of real-time adaptive learning mechanisms is considered an inevitable direction, helping the system to promptly detect emerging fraudulent behaviors and respond quickly to changes in the transaction environment. In the future, research should continue to focus on developing highly explainable ensemble models, improving methods for handling severely imbalanced data, and building adaptive AI-based analytical frameworks to improve fraud detection efficiency, while ensuring transparency, compliance, and sustainability across the entire digital banking ecosystem.

5. CONCLUSION AND RECOMMENDATIONS

This study has systematized and comprehensively analyzed scientific works related to the application of Decision Tree models in detecting digital banking transaction fraud. The overview results show that, despite its early appearance, Decision Trees still play a fundamental role in fraud detection systems thanks to their simplicity, clear and transparent interpretation in decision making, helping financial institutions understand and explain the operating mechanism of the model in the process of risk management and regulatory compliance. However, in the context of increasingly large, nonlinear, and rapidly fluctuating digital transaction data, traditional DT models need to be optimized in structure, adjusted in parameters, and combined with modern machine learning techniques to ensure forecasting performance, stability, and adaptability in real environments.

The overview results show that current research directions focus mainly on three main development streams.

First, detecting and classifying fraudulent transactions using the Single Decision Tree model or its variants, to ensure transparency and explainability in risk classification problems.

Second, addressing the problem of data imbalance – a common challenge in this field – using techniques such as SMOTE and its variants, which help improve the ability to identify minority classes and increase the sensitivity of the model.

Third, optimizing the Decision Tree model by integrating advanced machine learning methods such as Random Forest, Gradient Boosting, XGBoost, or cost-sensitive learning mechanisms and optimizing parameters using Grid Search or Bayesian Optimization. These research directions have contributed significantly to improving the accuracy, stability, and generalization ability of the model, thereby expanding its applicability in the field of digital banking. However, there are still significant research gaps regarding the ability to interpret results in complex models, computational efficiency when handling large-scale data, as well as real-time adaptability to monitor emerging fraudulent behaviors.

Based on that, this study proposes some future development directions. First of all, it is necessary to focus on developing hybrid Decision Tree models that are capable of self-learning and dynamic adaptation, allowing the system to update when new fraud patterns appear continuously. At the same time, integrating model explainability techniques (XAI) such as SHAP or LIME is necessary to ensure a balance between accuracy and transparency, making the model both strong in performance and reliable in monitoring and auditing. In addition, future studies should enhance the integration of real-time data combined with user behavior analysis, thereby building an early warning system capable of detecting potential fraud patterns before causing significant damage. Finally, from a governance perspective, the adoption of transparent, explainable, and highly adaptable machine learning models not only improves fraud detection efficiency but also strengthens risk management capabilities, strengthens compliance, and builds customer trust, contributing to shaping a more secure, sustainable, and transparent digital banking ecosystem in the future.

REFERENCES

1. P. Vanini, S. Rossi, E. Zvizdic, T. Domenig. Online payment fraud: from anomaly detection to risk management. *Financial Innovation*, **2023**, *9*(1), 66.

2. S. S. Nair, G. Lakshmikanthan, N. Belagalla, S. Belagalla, S. K. Ahmad, S. A. Farooqi. *Leveraging AI and machine learning for enhanced fraud detection in digital banking system: a comparative study*, The 1st International Conference on Advances in Computer Science, Electrical, Electronics, and Communication Technologies, Graphic Era Hill University, India, **2025**

3. PwC India. *Combating payments fraud in India's digital payments landscape*, PricewaterhouseCoopers Private Limited, New Delhi, 2025.

4. Bank for International Settlements. *Digital fraud and banking: Supervisory and financial stability implications*, BIS, Basel, **2023**.

5. Deloitte Center for Financial Services. *2024 Financial Services Industry Predictions*, Deloitte Insights, New York, **2024**.

6. W. Wei, J. Li, L. Cao, Y. Ou, J. Chen. Effective detection of sophisticated online banking fraud on extremely imbalanced data. *World Wide Web*, **2013**, *16*(4), 449-475.

7. S. K. Hashemi, S. L. Mirtaheri, S. Greco. Fraud detection in banking data by machine learning techniques. *IEEE Access*, **2022**, *11*, 3034-3043.

8. J. R. Quinlan. Induction of decision trees. *Machine Learning*, **1986**, *1*(1), 81-106.

9. S. Bhattacharyya, S. Jha, K. Tharakunnel, J. C. Westland. Data mining for credit card fraud: A comparative study. *Decision Support Systems*, **2011**, *50*(3), 602-613.

10. M. J. Page, J. E. McKenzie, P. M. Bossuyt, I. Boutron, T. C. Hoffmann, C. D. Mulrow, L. Shamseer, J. M. Tetzlaff, E. A. Akl, S. E. Brennan, *et al*. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, **2021**, *372*, n71.

11. M. Carminati, R. Caron, F. Maggi, I. Epifani, S. Zanero. BankSealer: A decision support system for online banking fraud analysis and investigation. *Computers & Security*, **2015**, *53*, 175-186.

12. E. A. Minastireanu, G. Mesnita. An analysis of the most used machine learning algorithms for online fraud detection. *Informatica Economica*, **2019**, *23*(1), 5-16.

13. S. Khatib. The Application of Machine Learning Models in Fraud Detection and Prevention Across Digital Banking Channels and Payment Platforms. *International Journal of Advanced Computational Methodologies and Emerging Technologies*, **2024**, *14*(9), 1-7.

14. H. Prabowo. Learning fraud detection from big data in online banking transactions: a systematic literature review. *Journal of Telecommunication, Electronic and Computer Engineering*, **2016**, *8*(3), 127-131.

15. D. A. Oduro, J. N. Okolo, A. D. Bello, A. T. Ajibade, A. M. Fatomi, T. S. Oyekola, và S. F. Owoo-Adebayo. *AI-powered fraud detection in digital banking: Enhancing security through machine learning*, *International Journal of Science and Research Archive*, **2025**, *14*(3), 1412-1420.

16. O. I. Odufisan, O. V. Abhulimen, E. O. Ogunti**.** Harnessing artificial intelligence and machine learning for fraud detection and prevention in Nigeria. *Journal of Economic Criminology*, **2025**, *7*, 100127.

17. H. Abbassi, A. Berkaoui, S. Elmendili, & Y. Gahi. End-to-end real-time architecture for fraud detection in online digital transactions. *International Journal of Advanced Computer Science and Applications*, **2023**, *14*(6), 749-757.

18. H. Abbassi, S. E. Mendili, Y. Gahi. Digital banking fortification: a real-time isolation forest architecture for detecting online transaction fraud. *Engineering Research Express*, **2024**, *6*(2), 025214.

19. K. Meduri. Cybersecurity threats in banking: Unsupervised fraud detection analysis. *International Journal of Science and Research Archive*, **2024**, *11*(2), 915-925.

20. S. Bhowmik, J. Howlader. *Online payment fraud monitoring and detection: Performance analysis of tree-based Ensemble Machine Learning models*, The 17th International Conference on Communication Systems and Networks (COMSNETS 2025), Bangalore, India, **2025**.

21. A. Ranjan, A. K. Jangir, S. S. K. Abrol. Online payment fraud detection using decision tree and LSTM neural network. *International Journal of Scientific Research in Engineering & Technology*, **2025,** *5*(5), 60-65.

22. A. A. Alhashmi, A. M. Alashjaee, A. A. Darem, A. F. Alanazi, R. Effghi. An ensemble-based fraud detection model for financial transaction cyber threat classification and countermeasures. *Engineering, Technology & Applied Science Research*, **2023**, *13*(6), 12433–12439.

23. D. Oluwadele, S. Malusi. *Performance analysis of ML algorithms for fraud detection in digital financial transactions*, The IEEE Conference on Information Communications Technology and Society, Capital Zimbali Conference Centre, Ballito, South Africa, **2025**

24. M. Marimuthu, K. Lekshmi, P. Saravanan, D. Nagaveni, P. Manikandan, B. Natarajan. *Transaction fraud detection using SMOTE oversampling*, The 1st International Conference on Software, Systems and Information Technology, Tumkur, India, **2024**.

25. P. Manisha, B. S. S. Kumar. *Enhanced NBA fraud detection in online banking using ensemble learning and SMOTE*, The 5th International Conference on Soft Computing for Security Applications, Tamil Nadu, India, **2025**.

26. N. Damanik, C. M. Liu. Advanced fraud detection: Leveraging K-SMOTEENN and stacking ensemble to tackle data imbalance and extract insights. *IEEE Access*, **2025**, *13*, 10356-10370.

27. E. M. Al-Dahasi, R. K. Alsheikh, F. A. Khan, G. Jeon. Optimizing fraud detection in financial transactions with machine learning and imbalance mitigation. *Expert Systems*, **2025**, *42*(2), e13682.

28. R. Padmavathy. *Cloud-based fraud detection and prevention solutions: using AI and machine learning to safeguard online banking transactions in the digital age*, *International Journal of Innovative Research in Computer and Communication Engineering*, **2021**, *9*(3), 102-110.

29. N. Hanbali, A. El-Yahyaoui. Advanced machine learning and deep learning approaches for fraud detection in mobile money transactions. *Innovations in Systems and Software Engineering*, **2025**, *21*(1), 1-21.

30. S. Kishan, K. Alluru. *Fraud detection in banking transactions using ensemble learning*, The 5th IEEE International Conference on Advances in Electronics, Computers and Communications, Bengaluru, India, **2023**.

31. B. Xu, Y. Wang, X. Liao, K. Wang. Efficient fraud detection using deep boosting decision trees. *Decision Support Systems*, **2023**, *175*, 114037.

32. M. A. Talukder, M. Khalid, M. A. Uddin. An integrated multistage ensemble machine learning model for fraudulent transaction detection. *Journal of Big Data*, **2024**, *11*(1),168.

33. G. Yu, Z. Luo. Financial fraud detection using a hybrid deep belief network and quantum optimization approach. *Discover Applied Sciences*, **2025**, *7*(5), 454.

34. Y. Salunke, S. Phalke, M. Madavi, P. Kumre, G. Bobhate, M. D. Madavi, P. D. Kumre. Fraud detection: a hybrid approach with logistic regression, decision tree, and random forest. *Cureus Journals*, **2025**, *2*(1), es44389-024-02350-5.

35. E. Btoush, X. Zhou, R. Gururajan, K. C. Chan, O. Alsodi. Achieving excellence in cyber fraud detection: A hybrid ML+DL ensemble approach for credit cards. *Applied Sciences*, **2025**, *15*(3), 1081.

36. D. Vijayanand, G. S. Smrithy. Explainable AI-enhanced ensemble learning for financial fraud detection in mobile money transactions. *Intelligent Decision Technologies*, **2025**, *19*(1), 52-67.